UNITED STATES ENVIRONMENTAL PROTECTION AGENCY

WASHINGTON, DC 20460

October 2, 2002

OFFICE OF
ENVIRONMENTAL INFORMATION

MEMORANDUM

SUBJECT: Peer Review of *Guidance on Geospatial Data Quality Assurance Project Plans (EPA QA/G-5G)*

FROM: Nancy W. Wentworth */s/ Nancy W. Wentworth*
Director, Quality Staff (2811R)

TO: Peer Review Panel

Attached is the October 2002, Peer Review Draft of *Guidance on Geospatial Data Quality Assurance Project Plans (EPA QA/G-5G)*. This technical guidance was developed by EPA to assist those involved in developing Quality Assurance (QA) Project Plans for geospatial data projects in accordance with *EPA Requirements for QA Project Plans (EPA QA/R-5)*. It discusses issues to be addressed in the QA Project Plan elements in the context of data collection and use of geospatial data, emphasizing systematic planning, the use of existing data, and hardware and software configuration issues for geographic information systems, and the graded approach.

This Peer Review version of the guidance incorporates significant revisions to the initial draft developed by the Geospatial Quality Council (dated January 2001), including changes made in response to the peer input review, conversion to Plain English, and updated examples of the graded approach.

The document contains four chapters. The introduction to the document provides an overview for the target audience emphasizing the importance of systematic planning for geospatial data projects and its relationship to the Agency's Quality System. The second chapter delineates the EPA policies on QA Project Plans. The third chapter discusses how QA Project Plan elements can be applied to geospatial data projects. The fourth chapter describes approaches to QA Project Plan elements for a geospatial module portion of a complex multi-organization project in risk assessment modeling as well as less complicated projects for a routine global positioning survey and creating a cartographic product from a spreadsheet to illustrate application of the graded approach discussed in Chapter 1.

You are asked to review all aspects of the document for relevance, usefulness, and overall adequacy as guidance for preparing QA Project Plans for geospatial data projects and to provide peer review comments. Your overall review is most appreciated, as well as your comments on the following questions:

1. Does the document begin with a clear indication of what it aims to address and how it would benefit the user?

2. Is it clear this is guidance? The document attempts to avoid being prescriptive, but rather to describe a variety of issues to be addressed depending on the nature of the project and its product's intended use. Is this the right message, and is it communicated effectively?

3. The document discusses both data collection and use activities. Does it strike the right balance between them in terms of relative emphasis and level of detail?

4. Are the examples helpful in demonstrating the principle of the graded approach?

5. Overall, how useful will this guidance be for its intended audience?

Please feel free to offer comments and suggestions that go beyond this charge, as you see fit. Use the line numbers provided in the document to reference specific sections with recommended changes. I appreciate your assistance in this review and would appreciate your comments by December 6, 2002. Please send written comments to:

Linda Kirkland (2811R)
U.S. Environmental Protection Agency
1200 Pennsylvania Avenue, N.W.
Washington, D.C. 20460
Phone: (202) 564-6873
Fax: (202) 565-2441
E-mail: kirkland.linda@epa.gov

Attachment

Guidance for Geospatial Data Quality Assurance Project Plans

EPA QA/G-5G

Quality Staff
Office of Environmental Information
United States Environmental Protection Agency

Washington, DC 20460

PEER REVIEW DRAFT

OCTOBER 2002

FOREWORD

The U.S. Environmental Protection Agency (EPA) has developed the Quality Assurance (QA) Project Plan as a tool for project managers and planners to document the type and quality of data and information needed for making environmental decisions. This document, *Guidance for Geospatial Data Quality Assurance Project Plans (EPA QA/G-5G)*, contains advice and recommendations for developing a QA Project Plan for projects involving geospatial data, including both newly collected or acquired data from other sources.

This document was designed for internal use and provides guidance to EPA program managers and planning teams. It does not impose legally binding requirements and may not apply to a particular situation based on the circumstances. EPA retains the discretion to adopt approaches on a case-by-case basis that differ from this guidance where appropriate. EPA may periodically revise this guidance without public notice.

This document is one of the U.S. EPA Quality System Series documents. These documents describe the EPA policies and procedures for planning, implementing, and assessing the effectiveness of the Quality System. As required by EPA Order 5360 A1 (EPA, 2000a), this document is valid for a period of up to five years from the official date of publication. After five years, this document will be reissued without change, revised, or withdrawn from the U.S. EPA Quality System Series. Questions regarding this document or other Quality System Series documents should be directed to the Quality Staff at:

U.S. EPA
Quality Staff (2811R)
1200 Pennsylvania Ave., NW
Washington, DC 20460
Phone: (202) 564-6830
Fax: (202) 565-2441
E-mail: quality@epa.gov

Copies of EPA Quality System Series documents may be obtained from the Quality Staff directly or by downloading them from their home page:

www.epa.gov/quality

ACKNOWLEDGMENTS

This document reflects efforts to adapt the QA Project Plan elements (EPA, 2001b) to projects involving geospatial data collection and use. The contribution of the Geospatial Quality Council members to the first draft of this document, and subsequent input from selected geospatial data users and the Quality Staff to this Peer Review Draft report, is greatly appreciated.

TABLE OF CONTENTS

			Page
CHAPTER 1	**INTRODUCTION**		**1**
1.1	WHAT IS THE PURPOSE OF THIS DOCUMENT?		1
1.2	WHY IS PLANNING FOR GEOSPATIAL PROJECTS IMPORTANT?		2
1.3	WHAT IS EPA'S QUALITY SYSTEM?		4
1.4	WHAT QUESTIONS WILL THIS GUIDANCE HELP TO ADDRESS?		6
1.5	WHO CAN BENEFIT FROM THIS DOCUMENT?		6
CHAPTER 2	**OVERVIEW TO CREATING A QA PROJECT PLAN**		**9**
2.1	INTRODUCTION		9
2.2	RELATED QA PROJECT PLAN GUIDANCE AND DOCUMENTATION		10
2.3	QA PROJECT PLAN RESPONSIBILITIES		11
2.4	SECONDARY USE OF DATA		12
2.5	REVISIONS TO QA PROJECT PLANS		12
2.6	OVERVIEW OF THE COMPONENTS OF A QA PROJECT PLAN		13
CHAPTER 3	**GEOSPATIAL DATA QA PROJECT PLAN GROUPS AND ELEMENTS**		**15**
3.1	INTRODUCTION		15
	3.1.1 A1.	Title and Approval Sheet	15
	3.1.2 A2.	Table of Contents	15
	3.1.3 A3.	Distribution List	16
	3.1.4 A4.	Project/Task Organization	16
	3.1.5 A5.	Problem Definition/Background	19
	3.1.6 A6.	Project/Task Description	19
	3.1.7 A7.	Quality Objectives and Criteria	20
	3.1.8 A8.	Special Training/Certification	22
	3.1.9 A9.	Documents and Records	22
3.2	GROUP B:	DATA GENERATION AND ACQUISITION	23
	3.2.1 B1.	Sampling Process Design	24
	3.2.2 B2.	Sampling and Image Acquisition Methods	27
	3.2.3 B3.	Sample Handling and Custody	28
	3.2.4 B4.	Analytical Methods	30
	3.2.5 B5.	Quality Control	31
	3.2.6 B6.	Instrument/Equipment Testing, Inspection, and Maintenance	32
	3.2.7 B7.	Instrument/Equipment Calibration and Frequency	33

Page

3.2.8 B8. Inspection/Acceptance Requirements for Supplies and Consumables 34
3.2.9 B9. Data Acquisition Requirements (Nondirect Measurements) 35
3.2.10 B10. Data Management 38
3.3 GROUP C: ASSESSMENT/OVERSIGHT 50
3.3.1 C1. Assessments and Response Actions 50
3.3.2 C2. Reports to Management 56
3.4 GROUP D: DATA VALIDATION AND USABILITY 57
3.4.1 D1. Data Review, Verification, and Validation 57
3.4.2 D2. Verification and Validation Methods 59
3.4.3 D3. Reconciliation with User Requirements 60

CHAPTER 4 GRADED APPROACH EXAMPLES 63
4.1 MINIMUM DOCUMENTATION EXAMPLE: CREATING A CARTOGRAPHIC PRODUCT FROM A SPREADSHEET CONTAINING FACILITY LATITUDE/LONGITUDE COORDINATES 65
4.1.1 Group A: Project Management 65
4.1.2 Group B: Measurement/Data Acquisition 67
4.1.3 Group C: Assessment/Oversight 68
4.1.4 Group D: Data Validation and Usability 69
4.2 MEDIUM DOCUMENTATION EXAMPLE: ROUTINE GLOBAL POSITIONING SURVEY TASK TO PRODUCE A GIS DATA SET 70
4.2.1 Group A: Project Management and Systematic Planning to Define the Task 70
4.2.2 Group B: Data Collection 71
4.2.3 Group C: Assessment and Oversight 73
4.2.4 Group D: Data Validation and Usability 74
4.3 COMPLEX DOCUMENTATION EXAMPLE: DEVELOPING COMPLEX DATA SETS IN A GIS FOR USE IN RISK ASSESSMENT MODELS 74
4.3.1 Group A: Project Management 75
4.3.2 Group B: Measurement/Data Acquisition 77
4.3.3 Group C: Assessment/Oversight 78
4.3.4 Group D: Data Validation and Usability 80

APPENDIX A: BIBLIOGRAPHY A-1
APPENDIX B: GLOSSARY B-1
APPENDIX C: PRINCIPAL DATA QUALITY INDICATORS FOR GEOSPATIAL DATA C-1

LIST OF FIGURES

		Page
Figure 1.	The EPA Quality System Approach to Addressing Geospatial Data Applications	4
Figure 2.	Steps of the Systematic Planning Process	5
Figure 3.	An Example Table of Contents and Distribution List	17
Figure 4.	An Example Organizational Chart	18
Figure 5.	GIS Flow Diagram	39

LIST OF TABLES

		Page
Table 1.	Questions that this Guidance Will Help to Address	6
Table 2.	EPA QA Guidance Documents	14
Table 3.	Summary of QA Groups and Elements	46
Table 4.	Continuum of Geospatial Projects with Differing Intended Uses	64

LIST OF ACRONYMS

DQO	Data quality objectives
EPA	U.S. Environmental Protection Agency
FGDC	Federal Geographic Data Committee
GIS	Geographic information system
GPS	Global positioning system
QA	Quality assurance
QC	Quality control
RMSE	Root mean square error
SSURGO	Soil Survey Geographic (data produced by the U.S. Natural Resources Conservation Service)
TIGER	Topologically Integrated Geographic Encoding and Referencing

CHAPTER 1

INTRODUCTION

Quality Assurance Project Plan: "A document describing in comprehensive detail the necessary Quality Assurance (QA), Quality Control (QC), and other technical activities that must be implemented to ensure that the results of the work performed will satisfy the stated performance criteria" *[EPA Requirements for Quality Assurance Project Plans (QA/R-5)* (EPA, 2001b, glossary)].

1.1 WHAT IS THE PURPOSE OF THIS DOCUMENT?

The EPA Quality System defined in EPA Order 5360.1 A2, *Policy and Program Requirements for the Mandatory Agency-wide Quality System* (EPA 2000d), includes coverage of environmental data or "any measurement or information that describe environmental processes, location, or conditions; ecological or health effects and consequences; or the performance of environmental technology. For EPA, environmental data includes information collected directly from measurements, produced from models, and compiled from other sources such as databases or literature." The EPA Quality System is based on an American National Standard, ANSI/ASQC E4-1994.

Consistent with the National Standard, E4-1994, Section §6.a.(7) of EPA Order 5360.1 A2 states that EPA organizations will develop a Quality System that includes "approved Quality Assurance (QA) Project Plans, or equivalent documents defined by the Quality Management Plan, for all applicable projects and tasks involving environmental data with review and approval having been made by the EPA QA Manager (or authorized representative defined in the Quality Management Plan). More information on EPA's policies for QA Project Plans are provided in Chapter 5 of the EPA Manual 5360 A1, *EPA Quality Manual for Environmental Programs* (EPA, 2000a) and *Requirements for Quality Assurance Project Plans (QA/R-5)* (EPA, 2001b). This guidance helps to implement the policies defined in Order 5360.1 A2. It is intended to help geospatial professionals who are unfamiliar with the requirements of QA Project Plans develop a document that meets EPA standards.

This guidance document describes the type of information that would be included in a QA Project Plan for a geospatial data project. Using this guidance, anyone from a geographic information system (GIS) technician at an EPA extramural supplier (e.g., contractor, university, or other organization) to an EPA Project Manager, Work Assignment Manager, or other EPA staff member, will know what information is needed in a QA Project Plan for projects involving geospatial data.

After reviewing this guidance document, the reader will have a clearer understanding of how to comply with these policies for geospatial projects. Not all elements of a QA Project Plan [as described in EPA's *Guidance for Quality Assurance Project Plans (QA/G-5)* (EPA, 1998a)] are applicable to all geospatial projects. Therefore, this guidance is provided to assist in the development of a QA Project Plan that is appropriate for the project. The elements, as described in the general EPA guidance on QA Project Plans (EPA, 1998a), are written with a focus on environmental data collection. This guidance helps the reader interpret those requirements for a geospatial project.

This document is just one of many documents that support EPA's Quality System. Quality Management Plans and other EPA Quality System documents are not discussed in detail in this guidance, but are also relevant and applicable to the use of geospatial data for or by EPA. Several other related documents may also serve as useful references during the course of a project, especially when other types of environmental data are acquired or used. This geospatial guidance supplements the *Guidance on Quality Assurance Project Plans (QA/G-5)* (EPA, 1998a).

1.2 WHY IS PLANNING FOR GEOSPATIAL PROJECTS IMPORTANT?

Planning is important in geospatial projects because it allows the project team to identify potential problems that may be encountered on a project and develop ways to work around or solve those problems before they become critical to timelines, budgets, or final product quality. Many examples exist of how a lack of planning impacts quality in geospatial projects. Lack of planning and detailed knowledge about data needs can cost a project a great deal of time and effort. Also the "graded" approach to developing QA Project Plans increases efficiency in that QA Project Plan elements are planned to be commensurate with the scope, magnitude, or importance of the project itself (See discussion in Chapter 4).

Example: Importance of Planning

Consider the case in which planning was not conducted on a project that required existing geospatial soils data. The project team needed a good quality source of soils data in a geospatial format. They decided to use the Soil Survey Geographic (SSURGO) data produced by the U.S. Natural Resources Conservation Service. SSURGO provides highly detailed soils data and had the content they required. They began their project by downloading SSURGO data for a single pilot project area and developed a series of applications programs over several weeks to correctly analyze and process the SSURGO data. When they had completed the pilot successfully, they began downloading the SSURGO data for the remainder of the study areas throughout the country. Only then did they discover that SSURGO data were only available in certain parts of the country and did not cover two-thirds of their project sites. The project team had to choose a different soils database and re-engineer their entire project to make use of this different geospatial data set.

A good QA Project Plan is valuable to a geospatial project in the following ways:

- It can be used to guide project personnel through the development process, helping ensure that choices are consistent with the established objectives and requirements for the project.

- Because the document fully describes the plans for the project, it will lead to a project with more transparency, better communication among the project team members, and better results for the decision maker.

- Using a QA Project Plan reduces the risk of schedule and budget overruns.

- If the QA Project Plan is properly followed, the project will lead to a more defensible outcome than a project without proper planning documentation.

- It will document the criteria and assumptions in one place for easy review and referral by anyone interested in the process.

- It uses a consistent format, making it easy for others to review the procedures and ensuring that individual steps are not overlooked in the planning phase.

In addition to these benefits, a project with a well-defined QA Project Plan often takes less time and effort to complete than a project without a planning document. Projects without planning documents are more likely to require additional cost and time to correct or redo collection, analysis, or processing of environmental data. The savings resulting from good planning typically outweighs the time and effort spent to develop the QA Project Plan. Poor quality planning often results in poor decisions. The costs of decision-making mistakes can be enormous and far outweigh the costs of proper planning for quality.

<u>What are the characteristics of a scientifically sound geospatial data project plan</u>? A scientifically sound, quality-based geospatial QA Project Plan:

- provides documentation of the outcome of the systematic planning process;
- is developed using a process designed to minimize errors;
- documents the standard operating procedures that will be followed;
- documents the data sources, format, and status of the existing data to be used in the project [including topological status, accuracy, completeness, and other required Federal Geographic Data Committee (FGDC) metadata];
- is frequently updated as new information becomes available or as changes in methodology are requested; and

- provides for the documentation of any changes from the original plan.

1.3 WHAT IS EPA'S QUALITY SYSTEM?

EPA has developed comprehensive requirements and procedures to include QC and QA in the planning stage of every project involving the use of environmental data. The EPA Quality System is described in EPA Order 5360.1 A2 (EPA, 2000d), which contains policy and program requirements for the mandatory, Agency-wide quality system. Emphasis is placed on planning for quality in projects before they have begun, rather than performing quality assurance and quality control planning during or after a project has been completed.

Figure 1 illustrates the role of a QA Project Plan for geospatial data projects within the context of the EPA Quality System. This guidance document describes all essential quality assurance information needed for a geospatial project. The figure shows the flow of data through data collection; data processing and analysis; and data validation, review, and assessment.

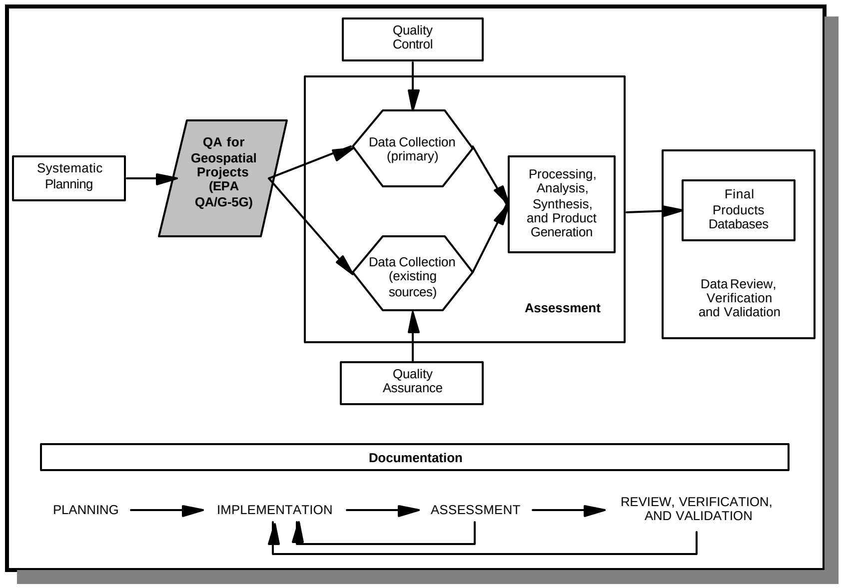

Figure 1. The EPA Quality System Approach to Addressing Geospatial Data Applications

The EPA Quality System is a management system that provides the elements necessary to plan, implement, document, and assess the effectiveness of QA and QC activities applied to environmental programs conducted by or for EPA. The EPA Quality System encompasses the collection, evaluation, and use of environmental data by or for EPA and the design, construction, and operation of environmental technology by or for EPA. EPA's Quality System has been built to ensure that environmental programs are supported by the type, quality, and quantity of data needed for their intended use. The EPA Quality System integrates policy and procedures, organizational responsibilities, and individual accountability.

How does systematic planning relate to a QA Project Plan? Systematic planning identifies the expected outcome of the project; its technical goals, cost, and schedule; and the criteria for determining whether the inputs and outputs of the various intermediate stages of the project, as well as the project's final product, are acceptable. The goal is to ensure that the project will produce the right type, quality, and quantity of data to meet the user's needs. EPA Order 5360.1 A2 (EPA, 2000d) requires projects for EPA environmental programs to use a systematic planning process to develop acceptance or performance criteria when collecting, evaluating, or using environmental data.

The systematic planning process can be applied to any type of data-generating project. The seven basic steps of the systematic planning process are illustrated in Figure 2. The first three steps can be considered preliminary aspects of scoping and defining the geospatial data collection or processing effort, while the last four steps relate closely to the establishment of performance criteria or acceptance criteria that will help ensure the quality of the project's outputs and conclusions. Performance and acceptance criteria are measures of data quality established for specific data quality indicators and used to assess the sufficiency of collected information. *Performance criteria* apply to information that is collected for the project. These criteria apply to new data. *Acceptance criteria* apply to the adequacy of existing information proposed for inclusion in the project. These criteria apply to data drawn from existing sources. Generally, performance criteria are used when data quality is under the project's control, while acceptance criteria focus on whether data generated outside the project are acceptable for their intended use on the project (e.g., as input to GIS processing software).

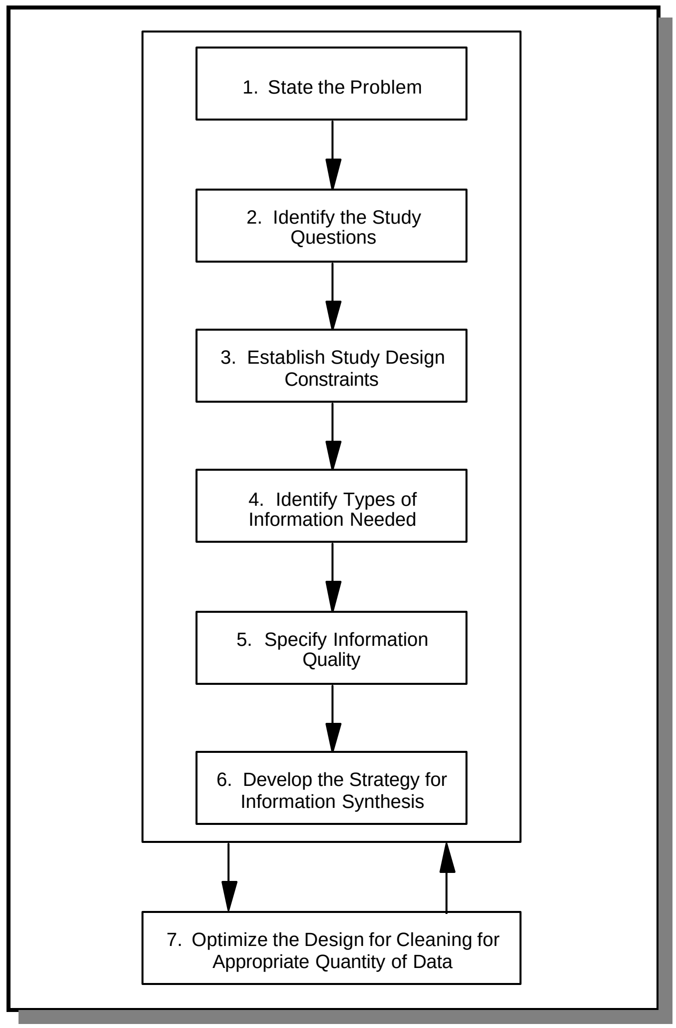

Figure 2. Steps of the Systematic Planning Process

Systematic planning is based on a common-sense, graded approach. This means that the extent of systematic planning and the approach to be taken match the general importance of the project and the intended use of the data. For example, when geospatial data processing is used to help generate data either for decision making (i.e., hypothesis testing) or for determining compliance with a standard, EPA recommends that the systematic planning process take the form of the Data Quality Objectives (DQO) Process that is explained in detail within *Guidance for the Data Quality Objectives Process (QA/G-4)* (EPA, 2000c).

1.4 WHAT QUESTIONS WILL THIS GUIDANCE HELP TO ADDRESS?

For quick reference to the information in this document, Table 1 provides a summary of the main questions addressed, indicating the chapter and sections containing this information.

Table 1. Questions that this Guidance Will Help to Address

Questions	Relevant Sections
How should the results of the planning phase for a geospatial data project be documented in a QA Project Plan?	3.1.7, 3.2.9
What quality assurance documentation is needed?	3.1.9
How do I document the acceptable level of uncertainty?	3.1.7, 3.2.9
What are some of the important metrics of quality for evaluating geospatial data (e.g., sensitivity analysis for GIS) and how can this information be used?	Appendix C
How do I conduct and document the data evaluation process?	3.3, 3.4
How do I assess the quality of geospatial data obtained from other sources (i.e., secondary use of existing data)?	3.2.9
What is needed to plan for data management (the process) and hardware/ software configuration?	3.2.10
How do I document changes from the planned process described in the QA Project Plan?	Chapter 2

1.5 WHO CAN BENEFIT FROM THIS DOCUMENT?

Anyone developing geospatial projects or using geospatial data for EPA will benefit from this document. This document will help in the creation of a QA Project Plan that specifically addresses the issues and concerns related to the quality of geospatial data, processing, and analysis. This document will help anyone who is:

- creating geospatial data from maps, aerial photos, or other sources;
- generating or acquiring the aerial photos;

- using existing data sources in their geospatial projects;
- generating new geospatial data from Global Positioning System (GPS) receivers;
- developing complex analysis programs that manipulate geospatial data;
- overseeing applications programming or software development projects—to understand how planning is related to developing software programs that use geospatial data;
- reviewing QA Project Plans for geospatial data—to understand the steps and details behind the planning;
- serving as a QA Officer for a group that creates or uses geospatial data.

CHAPTER 2

OVERVIEW TO CREATING A QA PROJECT PLAN

2.1 INTRODUCTION

> The QA Project Plan is the critical planning document for any environmental data collection operation because it documents how QA and QC activities will be implemented during the life cycle of a program, project, or task. The QA Project Plan is the blueprint for identifying how the quality system of the organization performing the work is reflected in a particular project and in associated technical goals (EPA, 1998a).

As explained in Chapter 1, QA Project Plans are necessary for all work performed by or for EPA that involves the acquisition of environmental data generated from direct measurement activities, collected from other sources, or compiled from computerized databases. This chapter provides more information on the source and intent of these policies and provides information on other related guidance and requirements documents, roles and responsibilities in creating QA Project Plans, and information on how and when to update QA Project Plans.

What is the purpose of a QA Project Plan? The QA Project Plan documents the systematic planning process for any data collection or use activity, as it documents how QA and QC activities will be planned and implemented. To be complete, the QA Project Plan will meet certain guidelines for detail and coverage (see *EPA Requirements for Quality Assurance Project Plans (QA/R-5)* (EPA, 2001b), but the extent of detail is dependent on the type of project, the data to be acquired and processed, the questions to be answered, and the decisions to be made. Overall, the QA Project Plan is to provide sufficient detail to demonstrate that:

- the project's technical and quality objectives are identified and agreed upon;
- the intended data acquisition and data processing methods are appropriate for achieving project objectives;
- the assessment procedures are sufficient for confirming that output data and products of the type and quality needed are obtained;
- any limitations on the use of the output data and products can be identified and documented.

EPA allows for flexibility in the organization and content of a QA Project Plan to meet the unique needs of each project or program. Although most QA Project Plans will describe project- or task-specific activities, there may be occasions when a *generic* QA Project Plan may be more appropriate. A generic QA Project Plan addresses the general, common activities of a program that are to be conducted at multiple locations or over a long period of time; for example, a large monitoring program that uses the same methodology at different locations. A generic QA Project Plan describes, in a single document, the information that is not site- or time-specific but applies throughout the program. Application-specific information is then added to the approved QA Project Plan as that information becomes known or completely defined. A generic QA Project Plan is reviewed periodically to ensure that its content continues to be valid and applicable to the program over time [*EPA Requirements for Quality Assurance Project Plans (QA/R-5)* (EPA, 2001b)].

2.2 RELATED QA PROJECT PLAN GUIDANCE AND DOCUMENTATION

Complex, broad-scope projects involving environmental data and geospatial databases may involve developing QA Project Plans that cross over many boundaries. For example, a multiyear, human health risk assessment project may involve taking and analyzing air samples from industrial sites, developing sophisticated software models, developing complex GIS procedures to process and analyze existing data from sources external to the project for use in the models, creation of new geospatial data, use of aerial photographs for ground-truthing,[1] and perhaps creating land-cover layers from new satellite imagery. Projects such as these may have more than one QA Project Plan. For example, there may be an overall QA Project Plan that establishes quality procedures, policies, and techniques for the project as a whole. Then for each subtask that contains a substantial amount of work or contains activities that in themselves require QA Project Plans, additional QA Project Plans may be required. In the example mentioned above, the following QA Project Plans would be needed:

- overall QA Project Plan that describes the quality system to be used on the project;
- QA Project Plan for the geospatial data aspects of the data collection and analysis;
- QA Project Plan for collection and analysis of air samples.

Each of these QA Project Plans may have similar information regarding overall project scope, purpose, management structure, and so on. But within the other QA groups—namely, Measurement and Data Acquisition (Group B), Assessment/Oversight (Group C), and Data Validation and Usability (Group D)—each QA Project Plan would contain specific and detailed information and procedures concerning the activities to be carried out for that specific project, be it environmental sampling, modeling development, or geospatial data use. Relevant documents can be found at www.epa.gov/quality.

[1] The use of a ground survey to confirm the findings of an aerial survey or to calibrate quantitative aerial or satellite observations.

2.3 QA PROJECT PLAN RESPONSIBILITIES

Who is responsible for creating a QA Project Plan? The QA Project Plan may be prepared by an in-house EPA organization (such as the GIS group), a contractor, an assistance agreement holder, or another federal agency under an interagency agreement. Most likely, the QA Project Plan will be a cooperative endeavor involving product users (e.g., EPA program managers funding the project), project managers responsible for the successful completion of the project, QA professionals, and technical staff responsible for carrying out the work.

For projects having limited scope, the QA Project Plan can be developed by a small team consisting of the product user, the EPA Project Manager, the project leader, and the technical staff. It is a guide to ensure that the quality of final products and resulting decisions meet criteria specified at the origination of the project.

Except where specifically delegated, all QA Project Plans prepared by non-EPA organizations are to be approved by EPA before they are implemented. It is Agency policy that the QA Project Plan be reviewed and approved by an authorized EPA reviewer to ensure that the document contains the appropriate content and level of detail. This may be the EPA Project Manager with the assistance and approval of the EPA QA Manager (EPA, 2001a, Sec. 2.5). The project leader and QA officer are to evaluate any changes to technical procedures before submitting new information to EPA.

All QA Project Plans are to be implemented as approved for the intended work. The organization performing the work is responsible for implementing the approved QA Project Plan and ensuring that all personnel involved in the work have copies of the approved QA Project Plan and all other necessary planning documents. These personnel are to understand the quality guidelines prior to the start of data generation activities (EPA, 2001a, Sec. 2.6).

Personnel developing and reviewing a geospatial data QA Project Plan are to have the proper experience and educational credentials to understand the relevant issues. The QA Project Plan is to be prepared such that external reviewers can understand the technical and quality issues associated with the project.

Discussions between the work managers and the technical staff are essential to creating a useful QA Project Plan. Management alone may not have an in-depth understanding of the complexity of geospatial data and its potential pitfalls. Geoprocessors may understand the data well but may not have enough background and scope information from management to determine the type, quantity, and quality of data required to meet the intended use. Only through an open quality planning process where all responsible parties meet to discuss quality goals and criteria can a useful QA Project Plan be developed.

2.4 SECONDARY USE OF DATA

Secondary Use of Data is the use of environmental data collected for other purposes or from other sources, including literature, industry surveys, compilations from computerized databases and information systems, and results from computerized or mathematical models of environmental processes and conditions.

In geospatial projects, use of existing data from a source external to the project is almost always required. When designing a project and, in turn, developing a QA Project Plan, the question of which GIS data sources to use is important. For example, in a project where elevation data are required, criteria for selecting appropriate elevation data are needed. Determining which source of digital evaluation model data (e.g., based on guidelines for scale, quality, and level of detail) is most appropriate for a project would require a dialog with management and technical staff to address the differences between available data sources in order to determine which source could produce a product adequate for its intended use. This decision-making process and the outcomes of the decisions are to be included in the QA Project Plan.

2.5 REVISIONS TO QA PROJECT PLANS

Because of the complex and diverse nature of environmental data operations, changes to project plans, methods, and objectives are often required. When a substantive change is warranted, the QA Project Plan is to be modified to reflect the change and is to be submitted for approval.

According to EPA policy, a revised QA Project Plan is to be reviewed and approved by the same authorities that performed the original review. Changed procedures may be implemented only after the revision has been approved. Changes to the technical procedures are to be evaluated by the EPA QA Manager and Project Manager to determine if they significantly affect the technical and quality objectives of the geospatial data project. If the procedural changes are determined to have significant effects, the QA Project Plan is to be revised and re-approved, and a revised copy is to be sent to all the persons on the distribution list. Only after the revision has been received and approved (at least verbally with written follow-up) by project personnel is the change to be implemented.

For programs or projects of longer duration, QA Project Plans need at least annual review to conform to EPA policy.

Refer to *Guidance for Quality Assurance Project Plans (QA/G-5)* (EPA, 1998a) and *EPA Requirements for Quality Assurance Project Plans (QA/R-5)* (EPA, 2001b) (*www.epa.gov/quality*) for additional information on how to handle QA Project Plan revisions.

2.6 OVERVIEW OF THE COMPONENTS OF A QA PROJECT PLAN

This section provides a list of the components of a QA Project Plan included in *EPA Requirements for Quality Assurance Project Plans (QA/R-5)* (EPA, 2001b). The components of a QA Project Plan are categorized into "groups" according to their function and "elements" within each group that define particular components of each group and form the organizational structure of the QA Project Plan. QA groups are lettered and QA elements are numbered.
The four groups are:

Group A. Project Management—The elements in this group address the basic area of project management, including the project history and objectives, roles and responsibilities of the participants, etc. These elements ensure that the project has a defined goal, that the participants understand the goal and the approach to be used, and that the planning outputs have been documented.

Group B. Data Generation and Acquisition—The elements in this group address all aspects of project design and implementation. Implementation of these elements ensure that appropriate methods for sampling, measurement and analysis, data collection or generation, data handling, and QC activities are employed and are properly documented.

Group C. Assessment and Oversight—The elements in this group address the activities for assessing the effectiveness of project implementation and associated QA and QC activities. The purpose of assessment is to ensure that the QA Project Plan is implemented as prescribed.

Group D. Data Validation and Usability—The elements in this group address the QA activities that occur after the data collection or generation phase of the project is completed. Implementation of these elements ensures that the data conform to the specified criteria, thus achieving the project objectives.

Table 2 is a complete list of the QA Project Plan groups and elements. Subsequent chapters of this document provide detailed information about the guidelines for sections of specific relevance to geospatial data projects. Some titles of the QA Project Plan elements, listed in Table 2, are slightly different in subsequent chapters to emphasize the application to geospatial data.

Table 2. Summary of QA Groups and Elements

Group	Element	Title
A	1	Title and Approval Sheet
	2	Table of Contents
	3	Distribution List
	4	Project/Task Organization
	5	Problem Definition/Background
	6	Project/Task Description
	7	Quality Objectives and Criteria
	8	Special Training/Certification
	9	Documents and Records
B	1	Sampling Process Design
	2	Sampling and Image Acquisition Methods
	3	Sample Handling and Custody
	4	Analytical Methods
	5	Quality Control
	6	Instrument/Equipment Testing, Inspection, and Maintenance
	7	Instrument/Equipment Calibration and Frequency
	8	Inspection/Acceptance Requirements for Supplies and Consumables
	9	Data Acquisition Requirements (Nondirect Measurements)
	10	Data Management
C	1	Assessments and Response Actions
	2	Reports to Management
D	1	Data Review, Verification, and Validation
	2	Verification and Validation Methods
	3	Reconciliation with User Requirements

CHAPTER 3

GEOSPATIAL DATA QA PROJECT PLAN GROUPS AND ELEMENTS

3.1 INTRODUCTION

The *EPA Requirements for Quality Assurance Project Plans (QA/R-5)* (EPA, 2001b) describes the elements EPA has specified for QA Project Plans. This guidance document provides specifics on how to develop these components for geospatial data projects, including suggested items to be included for each element. Each of the QA Project Plan elements that are specified in EPA (2001b) are listed below and are described here for application to a geospatial data project.

3.1.1 A1. Title and Approval Sheet

Suggested Content:
- Title of plan
- Name of organization
- Names, titles, and signatures of appropriate officials
- Approval dates

What is the purpose of this element? The purpose of the approval sheet is to enable officials to ensure that the quality planning process has been completed before significant amounts of work have been completed on the project and to document their approval of the QA Project Plan.

What type of information should be included in this element? The title sheet clearly denotes the title of the project, the project sponsor, and the name of the organization preparing the QA Project Plan. It includes any additional information on the title sheet that is necessary for the project (e.g., project number, contract number, additional organizations involved).

The approval sheet (which may or may not be a separate page) lists the names and signatures of the officials who are responsible for approving the QA Project Plan. The approving officials typically include the organization's technical Project Manager, the organization's QA Officer or Manager, the EPA (or other funding agency) Technical Project Manager/Project Officer, the EPA (or other funding agency) QA Officer or Manager, and other key staff, such as the task manager(s) and QA Officer(s) of the data to be used or collected for the project.

3.1.2 A2. Table of Contents

Suggested Content:
- Table of contents
- List of tables, figures, references, and appendices
- Document control format when required by EPA Project Manager

What is the purpose of this element? The table of contents provides an overall list of the contents of the document and enables the reader to quickly find specific information in the document.

What type of information should be included in this element? The table of contents lists all sections, tables, figures, references, and appendices contained in the QA Project Plan. The major headings for most QA Project Plans closely follow the list of required elements; as shown in Figure 3. While the exact format of the QA Project Plan does not have to follow the sequence given here, it is generally more convenient, and it provides a standard format for the QA Project Plan reviewer.

The table of contents of the QA Project Plan may include a document control component when required by the EPA Project Manager or QA Manager. This information would appear in the upper right-hand corner of each page of the QA Project Plan when the document control format is desired. The document control component, together with the distribution list (as described in Element A3), facilitates control of the document to help ensure that the most current version or draft of the QA Project Plan is in use by all project participants. Each revision of the QA Project Plan would have a different revision number and date.

3.1.3 A3. Distribution List

Suggested Content:
- Individuals and organizations to receive approved QA Project Plan
- Individuals and organizations responsible for implementation
- Individuals and organizations who

What is the purpose of this element? This element is used to ensure that all individuals who are to have copies of or provide input to the QA Project Plan receive a copy of the document.

What type of information should be included in this element? All the persons designated to receive copies of the QA Project Plan, and any planned future revisions, would be listed in the QA Project Plan. This list, together with the document control information, will help the Project Manager ensure that all key personnel in the implementation of the QA Project Plan have up-to-date copies of the plan. Note that the approved QA Project Plan can be delivered electronically.

3.1.4 A4. Project/Task Organization

Suggested Content:
- Identified roles and responsibilities
- Documentation of the QA Manager's independence of the unit generating the data
- The individual responsible for maintaining the official QA Project Plan is identified
- Organization chart showing lines of responsibility and communication
- List of outside external organizations and subcontractors in the organization chart

What is the purpose of this element? The purpose of this element is to provide EPA and other involved parties with a clear understanding of the role that each party plays in the investigation or study and to provide the lines of authority and reporting for the project.

CONTENTS

Section

List of Tables iv
List of Figures v
A Project Management 1
1 Project/Task Organization 1
2 Problem Definition/Background 3
3 Project/Task Description 4
4 Data Quality Objectives 7
4.1 Project Quality Objectives 7
4.2 Measurement Performance Criteria 8
5 Documentation and Records 10
B Measurement Data Acquisition 11
6 Sampling Process Design 11
7 Analytical Methods Requirements 13
7.1 Organics 13
7.2 Inorganics 14
7.3 Process Control Monitoring 15
8 Quality Control Requirements 16
8.1 Field QC Requirements 16
8.2 Laboratory QC Requirements 17
9 Instrument Calibration and Frequency 19
10 Data Acquisition Requirements 20
11 Data Management 22
C Assessment/Oversight 23
12 Assessment and Response Actions 23
12.1 Technical Systems Audits 23
12.2 Performance Evaluation Audits 23
13 Reports to Management 24
D Data Validation and Usability 24
14 Data Review, Validation, and Verification Requirements 24
15 Reconciliation with Data Quality Objectives 26
15.1 Assessment of Measurement Performance 26
15.2 Data Quality Assessment 27

Distribution List

N. Watson, EPA/ORD (Work Assignment Manager)*
B. Walker, EPA/ORD (QA Manager)

J. Warburg, State University (Principal Investigator)
T. Downs, State University (QA Officer)
G. Johnston, State University (Field Activities)
F. Haller, State University (Laboratory Activities)

B. O'Donnell, State University (Data Management)
E. Reynolds, ABC Laboratories (Subcontractor Laboratory)
P. Lafferton, ABC Laboratories (QA Manager Subcontractor Laboratory)

*indicates approving authority

Figure 3. An Example Table of Contents and Distribution List

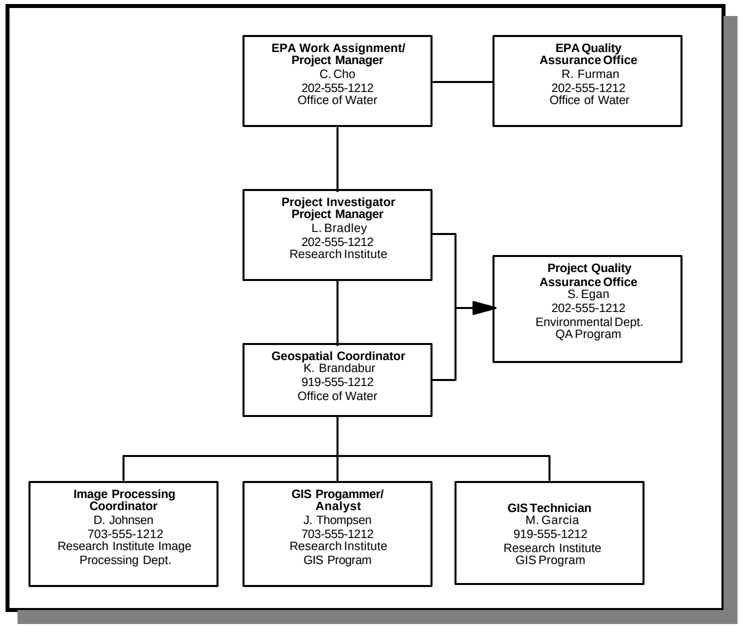

Figure 4. An Example Organizational Chart

<u>What type of information should be included in this element</u>? The specific roles, activities, and responsibilities of participants, as well as the internal lines of authority and communication within and between organizations, would be detailed. The position of the QA Manager or QA Officer would be described. The principal data users, decision maker, Project Manager, QA Manager, and all persons responsible for implementation of the QA Project Plan would be included—for example, data management personnel who maintain documentation of the initiation and completion of data searches, inquiries, orders, and order receipts, as well as of problems (e.g., incorrect or partial orders received, unacceptable overflights or film processing) and corrective actions that allow project managers to verify data acquisition progress. Also included would be the person responsible for maintaining the QA Project Plan and any individual approving deliverables other than the project manager. A concise chart showing the project organization, the lines of responsibility, and the lines of communication would be presented; an example is provided in Figure 4. For complex projects, it may be useful to include more than one chart—one for the overall project and others for each major subtask.

In geospatial projects for which GIS analysts acquire or collect geospatial data from external sources, the project organization element would describe how communications about these data (quality, completeness, problems acquiring, etc.) would be handled between the analyst and the project managers. The Project/Task Organization (A4) element designates individuals to whom staff can bring issues regarding project status and data quality. Additionally, it helps project managers know which technical staff will be responsible for performing each part of the project, better enabling management to obtain adequate status and quality information whenever necessary.

3.1.5 A5. Problem Definition/Background

Suggested Content:
- The specific problem to be solved or decision to be made
- Description of the project's purpose, goals, and objectives
- Identification of programs this project supports
- Description of the intended use of the data to be gathered

What is the purpose of this element? The purpose of this element is to describe the background and context driving the project and to identify and describe the problem to be solved or analyzed.

What type of information should be included in this element? The following types of information may be included:

- a description of the underlying purpose of the project;
- a description of the goals and objectives of the project;
- a description of the driving need for this project (e.g., regulation, legal directives, research, outreach);
- other projects, programs, or initiatives this project may be supporting;
- a description of the ultimate use of the final data or analysis;
- a description of the general overview of ideas to be considered and approaches to be taken on a particular project;
- the decision makers and/or those who will use the information obtained from the project.

3.1.6 A6. Project/Task Description

Suggested Content:
- Sufficient background for a historical and scientific perspective for project tasks
- Schedule and cost

What is the purpose of this element? The purpose of this element is to provide the participants with an understanding of the project tasks and the types of activities to be conducted. It includes a brief description of the data to be acquired and the associated quality goals, procedures, and timetables for project and task completion.

What type of information should be included in this element? Detailed descriptions of processing tasks will be created in Group B elements. Summaries and bulleted lists are adequate for most types of information to be included here. Items to consider including are:

- a description of the location of the study area and the processes and techniques that will be used to acquire necessary geospatial data;

- a description of any special personnel or equipment required for the specific type of work being planned;
- information on how data processing and management will be performed and by whom;
- identification and description of project milestones and the schedule associated with achieving these milestones;
- deliverables, the schedule associated with generating and submitting them, and the format to which these deliverables are to adhere;
- a work breakdown structure associated with the project, detailing the individual work components associated with the milestones and deliverables, whose progress will be tracked throughout the duration of the project.

3.1.7 A7. Quality Objectives and Criteria

> **Suggested Content:**
> - The quality objectives for the project
> - The performance and acceptance criteria used to evaluate quality. (Use the systematic planning process to develop quality objectives and performance criteria [see *EPA Quality Manual for Environmental Programs,* Section 3.3.8.1 (EPA, 2000a), for more

What is the purpose of this element? The purpose of this element is to document the quality objectives of the project and to detail performance and acceptance criteria through the systematic planning process that will be employed in generating the data. Performance and acceptance criteria can take many forms. The overall goal in setting the criteria is to ensure that the project will produce the right type, quality, and quantity of data to meet the user's needs.

Where does the information for this element come from? This information comes from the systematic planning process. The systematic planning process is a means of ensuring that the appropriate quality and quantity of data and processing are performed on the project to produce products adequate for their intended use. Systematic planning is required even when the project or task will not result in a definable decision. During systematic planning, performance criteria are to be specified so that, during quality assessment, there is a known benchmark against which quality can be gauged. The criteria for quality are to be set at a level commensurate with the project-specific requirements. In other words, performance and acceptance criteria specify the level of quality that would be acceptable for the final data or product. They are not to be set higher or lower than what is required to meet the needs of that particular project.

How are quality objectives and criteria determined? They are determined through the systematic planning process as the planning team reviews and discusses what is needed for the basic questions to be answered or the decision to be made with the project results (Section 1.3). For example, if a regulatory decision is the ultimate product of the task, then the Agency strongly recommends using the DQO Process. Data quality objectives are qualitative and quantitative statements that:

- clarify the intended use of the data;
- define the type of data needed to support the decision;
- identify the conditions under which the data are to be collected;
- specify tolerable limits on the probability of making a decision error due to uncertainty in the data.

For decision-making programs in which systematic planning takes the form of the DQO Process, these criteria are represented within data quality objectives (EPA, 2000b) that express data quality requirements to achieve desired levels of confidence in making decisions based on the data.

<u>What are some of the forms that performance or acceptance criteria might take in a geospatial data project</u>? Examples may include:

- a description of the resolution and accuracy required in input data sources;
- statements regarding the speed of applications programs written to perform data processing (e.g., "the programs must be able to make 10,000 Monte Carlo simulation runs within 8 hours");
- criteria for choosing among several existing data sources for a particular geospatial theme (e.g., land use); geospatial data needs are often expressed in terms of using the "best available" data, but different criteria—such as scale, content, time period represented, quality, and format—may need to be assessed to decide which are the "best available" (when more than one is available) to use on the project;
- specifications regarding the accuracy needs of coordinates collected from GPS receivers;
- requirements for aerial photography or satellite imagery geo-referencing quality, such as specifications as to how closely these data sources need to match spatially with ground-based reference points or coordinates;
- criteria to be met in ground-truthing classified satellite imagery.

If address geo-coding is to be performed, indicate the criteria for minimum overall match rate and any tolerances to be used in address matching procedures, including whether or not spatial offsets are to be supplied in the resulting coordinates and, if so, what the offset factor is to be. If the project is to build new geospatial data sets through a map digitizing process, indicate requirements for topology, label errors, attribute accuracy, overlaps and gaps, and other processing quality indicators.

Appendix C, *Principal Data Quality Indicators for Geospatial Data*, provides additional information regarding data quality indicators that could be reflected in quality criteria to be specified in this element.

3.1.8 A8. Special Training/Certification

Suggested Content:

- Any special training or certification requirements for the project
- Plans for meeting these requirements

What is the purpose of this element? The purpose of this element is to document any specialized training requirements necessary to complete the project. This element is a good place to discuss how these requirements will be met and how to verify that they have been met.

What type of information should be included in this element? Requirements for specialized training for field-sampling techniques such as global positioning technology, photo interpretation, and data processing would be specified. Depending on the nature of the project, the QA Project Plan may address compliance with specifically mandated training requirements (e.g., software contractors needing company certification or employees needing software training). This element of the QA Project Plan would show that the management and project teams are aware of specified health and safety needs as well as any other organizational safety plans. Training and certification for necessary personnel would be planned well in advance of the implementation of the project. All certificates or documentation representing completion of specialized training would be maintained in personnel files.

3.1.9 A9. Documents and Records

Suggested Content:

- Description of the mechanism for distributing the QA Project Plan to project staff
- List of the information to be included with final products, including metadata records, calibration and test results (for GPS or remote sensing tasks), processing descriptions provided by data vendors (e.g., address matching, success rate reports from address matching vendors)
- List of any other documents applicable to the project, such as hard-copy map source material, metadata provided with data from secondary data sources, interim reports, final reports
- All applicable requirements for the final disposition of records and documents, including location and length of retention period

What is the purpose of this element? This element defines which documents and records are critical to the project. It provides guidance to ensure that important documentation is collected, maintained, and managed so that others can properly evaluate project procedures and methods.

What type of information should be included in this element? This element could be used to provide guidelines for clearly documenting software programs (including revisions) and models, field operation records (for GPS activities), and metadata guidelines.

Metadata are required for geospatial data created on federal government contracts, and this element is a good place to indicate metadata requirements. Detailed metadata indicating the source, scale, resolution, accuracy, and completeness are needed to assess the adequacy of existing data for use (EPA, 2000d). The Federal Geographic Data Committee (*www.fgdc.gov*) has developed a metadata standard for geospatial data generated for and by all federal agencies which all federal agencies are required to follow according to Executive Order 12906, "Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure". If an external source of existing data does not supply metadata (preferably, Federal Geographic Data Committee–compliant metadata including quality data elements), or additional information from the external source cannot be obtained, then the quality of these data for this project cannot be evaluated. The data would be of unknown quality and unsuitable for producing a product adequate for its intended use.

Other types of documentation and records that would be described in the Documents and Records (A9) element include field operation records, analysis records, and data handling records. This element would be used to describe the generation of these records (where, by whom, and what format they will be stored and reported in). This element would discuss how these various components will be assembled to represent a concise and accurate record of all activities affecting data quality.

In some environmental sampling projects, records and documentation that refer to geospatial data collection may be included in the environmental sample planning portion of a general QA Project Plan, rather than in a geospatial QA Project Plan. In these cases, the GPS records are associated with the environmental sampling in general, not with the geospatial data records and documentation. The Documentation and Records (A9) element of a geospatial QA Project Plan could then reference the GPS records requirements that are described in the environmental sampling QA Project Plan.

3.2 GROUP B: DATA GENERATION AND ACQUISITION

Geospatial projects may involve the creation of new geospatial data from field measurements (e.g., from GPS measurement, aerial photography, or satellite imagery) or may involve the acquisition and use of existing geospatial data originally created for some other use. The Group B elements of the QA Project Plan are used to:

- describe the quality assurance and quality control of the instruments, procedures, and methods used to create new geospatial data (the first eight elements);

- describe the methods of acquiring, assessing, and managing data from existing sources for the project [Data Acquisition Requirements (Nondirect Measurements) (A9) and Data Management (B10) elements].

While the first eight elements are often associated with the creation of new data from measurements, the Quality Control (B5) element may be used to outline and document quality control procedures used on certain existing data sources. For example, it could be used to document quality control procedures when map digitizing will be performed or when classified satellite imagery is to be assessed for quality via ground-truthing procedures.

Data Acquisition Requirements (Nondirect Measurements) (B9) and Data Management (B10) elements are often the most significant parts of the Group B elements in geospatial projects. This is because geospatial projects almost always involve the use of existing data sources from outside organizations (e.g., existing geospatial data products like Topologically Integrated Geographic Encoding and Referencing data, Digital Line Graph data, National Land Cover Data, and Digital Elevation Model data). In addition, geospatial projects inherently involve data management—therefore, the Data Management (B10) element will require extensive inputs to the QA Project Plan since it is used to describe the data management procedures used to ensure that data are processed and handled in ways that meet the accuracy and quality required on the project.

Whereas the methods described in the Group B elements are *summarized* in the Project/Task Description (A6) element, the purpose here is to provide *detailed* information on the data collection procedures and methods.

3.2.1 B1. Sampling Process Design

<u>What is the purpose of this element</u>? This element describes all the relevant components of the data collection or image acquisition design, defines the key parameters to be estimated, and indicates the number and type of samples or images expected. It also describes where, when, and how samples or images are to be taken. The information is to be sufficiently detailed to enable a person knowledgeable in this area to understand how and why the samples or images will be collected. Most of this information may be available as outputs from the final steps of the systematic planning process.

Suggested Content:
- Description of the data acquisition design
- For existing data from other sources, how data will be evaluated for use
- For geospatial data to be collected, the design for acquisition (e.g., how and where locational data will be acquired)

What type of information should be included in this element? This element would be used to describe how the project will acquire the "right" data for the project. For example, if the project will be using satellite imagery, it is important to consider the type, quality, and resolution of imagery.

Use the Sampling Process Design (B1) element to describe the geographic extent of locational data to be acquired. Describe the size, shape, and location of the project's study area. Document whether it is feasible to collect new geospatial data for the project and why. If the project involves a number of discrete study areas (for example, a set of regulated industrial sites), data of differing dates, quality, resolution, or scale may be available. Determine whether different resolutions of data may be used in different parts of the project. This issue arises when very accurate data exist for some portions of a study, but not for others. An example issue to address in this element would be whether a single, uniform data source would be acceptable even though in some areas it does not contain the most recent data, or in some areas, the resolution is not as high as in the other data sources, would be addressed in this element.

When acquiring locational information using GPS equipment, this element would be used to describe the locations to be used and the rationale for this design. In many cases, GPS will be used to gather information at specific, known locations. For example, this element may specify that GPS data will be collected at each spotted owl nest found or at each outfall encountered along a body of water. For other projects, a sampling design may be implemented to collect data using sophisticated sampling techniques. For example, when collecting soil samples to be analyzed for contamination, sampling techniques may be used to determine the number of samples to be taken and the method for determining the locations (e.g., based on a systematic grid of predefined size or by using judgmental sampling procedures, etc.) The Sampling Process Design (B1) element would be used to describe the sampling design as it relates to the locations. The sampling design might take into account procedures for dealing with locally interfering objects such as tree canopies, towers, buildings, or high-relief terrain that could impact or eclipse the GPS signal. Within the description of the sampling design, this element would also describe the frequency of locational sampling or image acquisition. When decisions are made on the number and location of observations or images to be taken, the QA Project Plan would describe how these decisions were derived to meet the requirements of the planned interpretation (e.g., accuracy and precision requirements) or analysis.

Finally, the objectives for collecting the identified geospatial data are to be formulated in the planning stage of the project. This element would explain why these data are being acquired and how they will be used on the project.

What are some examples of issues that would be addressed in this element? Acquiring locational data with GPS frequently involves a certain amount of uncertainty regarding the exact location to be captured. This uncertainty can occur when collecting data for use in regulatory analyses. Some examples of the types of questions that could be addressed here include the following:

- When collecting industrial site information, what, precisely, is to be collected: the location of the facility main gate or the main office front door? the location of holding ponds or other waste units?

- Is it necessary to collect all waste unit locations or just the location of the general center of all the waste units?

- How important is the accuracy of these particular locations?

The Sampling Process Design (B1) element might also describe the frequencies and logistics involved in the GPS or imagery acquisition tasks. For example, information in this element would provide answers to questions such as:

- When do the data need to be collected, processed, and ready to be used on the project?

- Are there any constraints due to seasonality? For example, is imagery to be acquired with "leaf off" or "leaf on?" Can GPS acquisition be done on weekends?

- When performing work with plants and animals, what seasonal factors will affect the ability to find or track these species?

- What logistical activities can be planned to facilitate GPS data collection?

- Are special vehicles required?

- Will the sampling take place on water? If so, what provisions for water transportation are necessary?

To address some of these issues, the use of bar charts showing time frames of various QA Project Plan activities is recommended to identify both potential bottlenecks and the need for concurrent activities. The most appropriate plan for a particular direct measurement or remote sensing task will depend on the practicality and feasibility of the plan, the key characteristic to be estimated, and the resources needed for implementation (e.g., the costs of direct measurement or remote sensing and interpretation).

The Sampling Process Design (B1) element is the place to discuss the need for base station data, if applicable. In addition, for projects involving digitizing source maps directly into GIS format, issues related to evaluating source materials might be discussed.

What might be included in this element for projects involving acquisition of new aerial photography? This element would include issues related to precision, seasonality, resolution (pixel size), geo-registration techniques and quality, delivery medium (analog photos or digital orthophotography), and types and levels of vendor processing. An imagery acquisition plan could be used to identify the types of data required, spatial resolution, overpass date(s)/time(s), and supporting data required. Consider the following specific issues:

- What final surface characteristic(s) does the project require (e.g., vegetation type, canopy cover, soil type, or vegetation stress)? This derived parameter or analysis will determine what type of imagery is needed.

- For film-product aerial photography, are black-and-white, true-color, or false-color products needed?

- Is a particular time of year appropriate for imagery acquisition?

- What time of day are aerial photos or satellite images to be captured (usually not an option for satellite imagery, but may be for aerial photography)?

- What documentation is needed on climatic factors, such as maximum allowable cloud cover and snow cover?

3.2.2 B2. Sampling and Image Acquisition Methods

Suggested Content:
- Description of data collection procedures
- Methods and equipment to be used
- Description of GPS equipment preparation requirements
- Description of performance requirements
- Description of corrective actions to be taken if problems arise

What is the purpose of this element? This element would be used to document procedures and methods for collecting samples. As with all other considerations involving geospatial sampling or image acquisition, methods are to be chosen with respect to the intended application of the data. Different sampling or imagery acquisition methods have different operational characteristics—such as cost, difficulty, and necessary equipment—that affect the representativeness, comparability, accuracy, and precision of the final result.

What type of information should be included in this element? Consider systematic planning requirements when choosing the methods to ensure that : (1) the measurements, observations, or images accurately represent the portion of the environment to be characterized; (2) the locational coordinates sampled are of sufficient accuracy to support the planned data analysis; and (3) the locational coordinates sampled meet completeness requirements. Be sure that data collected via GPS will meet the requirements for the intended use. Use standard operating procedures to ensure that acquisition procedures are consistent across multiple staff members and that Agency standards are used when available.

Identify the type of direct measurement, observation, or image to be acquired and the appropriate sampling methods to be used from applicable methods approved by EPA. Each direct measurement, observation, or image has its own characteristics that define the method performance and the required sampling to represent the environment. Address the following:

- actual sampling locations
- choice of measurement or remote-sensing method
- delineation of a proper measurement, observation, or image entity
- inclusion of all entities within the abstract universe sampled (Appendix C addresses the need for completeness indicators).

This element would address the issues of responsibility for the quality of the data, the methods for making changes and corrections, the criteria for deciding on a new sample location, and documentation of these changes. It would describe appropriate corrective actions to take if there are serious flaws in the implementation of the sampling methodology. For example, if part of the complete set of GPS measurements or imagery samples to be acquired is found to be inadequate for its intended use, describe how replacements will be obtained and how these new samples will be integrated into the total set of data.

3.2.3 B3. Sample Handling and Custody

Suggested Content:
- Description of requirements for handling and transfer of hard-copy imagery or other hard-copy data inputs

What is the purpose of this element? This element is used to define the project-specific requirements for handling samples and, perhaps, hard-copy aerial photographs or other source documents such as maps. These project-specific requirements may be necessary to prove that source materials and samples have been properly handled and managed during the course of the project.

What type of information should be included in this element? Aerial photography delivered in hard-copy format may need to go through a chain-of-custody procedure. However, GPS coordinates, satellite imagery, and digital orthophotography are usually delivered and processed in electronic form. Therefore, the Sample Handling and Custody (B3) element has limited applicability on geospatial projects. The procedures for handling, maintaining, and processing electronic data are described in the

Data Acquisition Requirements (Nondirect Measurements) (B9) and Data Management (B10) elements.

Hard-copy aerial photography, original source maps, and hard copies of satellite imagery can sometimes be of great importance in geospatial projects. They may provide the only source of concrete information regarding industrial facilities and their surroundings, especially when historical aerial photos are available for particular areas. Therefore, these documents need to undergo careful and deliberate chain-of-custody procedures to ensure that they are not lost, misplaced, altered, or destroyed. This element is used to document chain-of-custody procedures and, for geospatial projects, may only be applicable for the QA Project Plan if hard-copy documents such as aerial photos are acquired and used. However, chain-of-custody procedures for environmental media samples (air, water, soil) would be developed and documented in QA Project Plans for the environmental sampling portions of the project.

For aerial photographs, source maps, and other hard-copy documents, this element is used to ensure that the documents are:

- transferred, stored, and analyzed by authorized personnel;
- not physically degraded through handling;
- properly recorded and tracked to ensure that their whereabouts are known at all times in case they need to be used by different researchers.

The QA Project Plan discusses the source material or imagery handling and custody procedure at a level commensurate with the intended use of the data. This discussion might include the following:

- a list of names and responsibilities of those who will be handling the documents;
- a description and example of the document numbering system;
- procedures that will be used to maintain the chain of custody and documentation of handling procedures within the organizations using these documents;
- the physical location and filing system to be used to store and manage the documents.

Few geospatial projects will need to fully develop a chain-of-custody process for source documents. However, for projects that do acquire and use rare, original, or irreplaceable source documents (aerial photos, printed maps, archival satellite imagery), it is a good idea to design and document chain-of-custody procedures.

The forms and procedures used to track the chain of custody of source documents could be described in the Documents and Records (A9) element. In this way, the documentation to be maintained would be described in Documents and Records (A9) element and the procedures themselves would be described in the Sample Handling and Custody (B3) element.

3.2.4 B4. Analytical Methods

Suggested Content:

- Image processing and/or photo-interpretation methods to be used
- List of method performance standards, if applicable

What is the purpose of this element? When GPS coordinates, aerial photos, or satellite imagery is to be processed or interpreted, the Analytical Methods (B4) element would be used to document these interpretation or processing methods. For remote sensing data sets, requirements may need to be developed for the image analysis or processing to produce new data sets. Image analysis may range from manual interpretation/characterization to the application of algorithms and/or models.

What type of information should be included in this element? This element would document algorithms/models to ensure they are applied correctly and consistently. For example, when using remote sensing data sets, some requirements may need to be developed for the image analysis or data processing that produces new data sets. Examples of new data sets derived from remote sensing are:

- plant biomass indices that convert visible and near infrared to a scalar value representing the relative amount of green vegetation

- land-cover classifications that segment an image into classes (pavement, water, vegetation) based on reflectance and/or thermal radiance of each pixel.

This element would address methods to be used, and in particular, whether the selected methods differ from standard procedures. For example, most biomass estimators such as the Normalized Difference Vegetation Index were developed to be applied to surface reflectance, not digital numbers or radiance values. If a conversion to reflectance is not performed, some justification would be noted. Statistics-based clustering (classification) of an individual image can be performed on the digital number values; however, if the classification is to be performed on multiple images, some type of image normalization would need to be performed. This element of the report would describe the approach used.

Similarly, for aerial photo interpretation tasks, the methods used to interpret the photos would be documented in this element. Existing standard operating procedures could be cited or included to describe the interpretation methods and relate them to the desired products to be generated from the interpretation.

3.2.5 B5. Quality Control

What is the purpose of this element? Quality control is the "overall system of technical activities that measures the attributes and performance of a process, item, or service against defined standards to verify that they meet the stated requirements established by the customer, operational techniques, and activities that are used to fulfill requirements for quality" (EPA, 2001b). The Quality Control (B5) element documents any QC checks not defined in other QA Project Plan elements and would reference other elements that contain this information, where possible. This element relies on performance criteria described in the Quality Objectives and Criteria (A7) element. In other words, use the Quality Objectives and Criteria (A7) element to describe acceptable performance criteria and use the Quality Control (B5) element to describe the procedures to be used to assess the performance.

What type of information would be included in this element? The Quality Control (B5) element is primarily applicable when generating new data, such as using GPS to collect coordinates, using a digitizing procedure to convert source maps into GIS formats, or using ground-truthing procedures to assess the accuracy of classified satellite imagery.

Suggested Content:
- QC activities needed for GPS measurements, field observations, map digitization, image acquisition, image processing, or image analysis
- The frequency of each check and corrective action required when limits are exceeded

QC checklists are often a means of ensuring that proper procedures are used at each step in data collection, or of checking and assessing the quality of map digitizing or satellite ground-truthing results. QC checklists could be developed and described in the Quality Control (B5) element to facilitate efficient and accurate fieldwork when using GPS receivers. QC checklists could help analysts and management ensure that equipment has been checked and is operating properly before fieldwork begins each day, and to ensure that proper procedures are used when collecting calibration points (first-order control points) as well as the coordinates themselves.

Including QC procedures to be used in map digitizing in the Quality Control (B5) element is important to ensure that digitizing staff convert the correct map features in a way that meets accuracy requirements. For example, describe checklists to be used by the digitizer to confirm that georegistration of the map-to-ground coordinates is within tolerances and that each required feature from the map is digitized and added to the appropriate GIS layer or feature class.

Quality control of classified satellite imagery would involve some ground-truthing procedures. These QC procedures may be documented in the Quality Control (B5) element and checklists to be completed by the responsible staff may be described.

What assessments would be done to verify that the criteria have been met? The assessment process includes verifying the data set (or product) specifications. The evaluations planned provide a basis for logical decisions on the applicability of the data or images to the current project. Examples include:

- ensuring that the requested special bands have been delivered;
- checking against independent data sets such as other images or vector products;
- examining the cloud coverage of images to ensure that cloud coverage extent does not impede use of the data;
- ensuring that the view angle of imagery is as specified.

Although the project-specific requirements listed above may seem rather simple, many geospatial projects have a large extent and variety of geospatial data. The directions in this element of the QA Project Plan ensure that all these data are evaluated systematically and completely.

The Quality Control (B5) element would also be used to document the actions to be taken if QC checks identify errors or failures in quality of data capture procedures.

3.2.6 B6. Instrument/Equipment Testing, Inspection, and Maintenance

Suggested Content:
- Description of how inspections and acceptance testing of instruments, equipment, and their components affecting quality will be performed and documented
- Description of how deficiencies will be resolved
- Description of (or reference to) periodic preventive and corrective maintenance of measurement or test

What is the purpose of this element? The purpose of this element is to discuss the procedures used to verify that all instruments and equipment are maintained in sound operating condition and are capable of operating at acceptable performance levels. This element provides a mechanism for ensuring that equipment used in geospatial projects is operating to specifications. If the project does not involve the use of any measurement equipment, then it can be stated that this element is not applicable in the QA Project Plan.

What type of information would be included in this element? Standard operating procedures may be referenced or included in the Instrument/Equipment Testing, Inspection, and Maintenance Requirements (B6) element to document the required procedures for equipment testing and inspection (e.g., for GPS equipment). Descriptions of procedures may include:

- estimates of the possible impact of equipment failure on overall data quality, including timely delivery of project results;

- any relevant site-specific effects (e.g., environmental conditions);
- steps for assessing the equipment status.

This element would address the scheduling of routine calibration and maintenance activities, the steps that will be taken to minimize instrument downtime, and the prescribed corrective actions for addressing unacceptable inspection or assessment results. This element would also include periodic maintenance procedures. Supply the reader with sufficient information to review the adequacy of the instrument/equipment management program.

Before a GPS survey is undertaken, it is recommended that equipment be tested to ensure that it works properly. Check the unit to confirm critical settings, because these settings remain in memory when the receiver is turned off; failure to do so could result in inaccurate data.

Routine preventive maintenance schedules need to be established and records maintained on all instruments, equipment, and computer hardware and software systems used for the acquisition of data, analysis of photographs, and graphics functions conducted. Designate appropriate personnel who use instruments and equipment requiring routine maintenance as responsible for ensuring that maintenance is performed in accordance with relevant standard operating procedures or equipment instructions, and that maintenance is properly documented. This will help ensure that maintenance records are available on request.

When aerial photography is needed in a geospatial project, inform the data producer of the requirement to provide documentation of the equipment used, as well as its maintenance and testing records, to assure project-specific requirements for their task are met.

3.2.7 B7. Instrument/Equipment Calibration and Frequency

Suggested Content:
- Instruments used for data collection whose accuracy and operation need to be maintained within specified limits
- Description of (or reference to) how calibration will be conducted
- How calibration records will be maintained and traced to the instrument

<u>What is the purpose of this element</u>? The purpose of this element is to identify the equipment to be calibrated and to document the calibration method and frequency of each instrument.

<u>What type of information might be included in this element</u>? Identify any equipment or instrument that requires calibration or standardization to maintain acceptable performance. Include or reference standard operating procedures that document how calibration of the equipment (e.g., for GPS receiver units) would be accomplished.

Generally, this will involve collecting locations with the GPS unit and comparing them to known, high-quality reference points.

Identify and describe the calibration or standardization method for each instrument in enough detail for someone else to duplicate the method. Reference external documents such as standard operating procedures, providing these documents can be easily obtained. Fully document and justify nonstandard methods.

If very high accuracy is required for locational data, geospatial data collectors can turn to reference calibration data supplied by National Institutes of Standards and Technology, which compares the frequency standard of each satellite to their frequency standard. (See *www.boulder.nist.gov/timefreq/service/gpstrace.htm*.)

Aerial photography firms might be requested to supply calibration documentation for the equipment used to capture any aerial photographs on the project. In addition, any film processing equipment calibration documentation (if receiving hard-copy photographs rather than electronic versions) would be included in this element.

3.2.8 B8. Inspection/Acceptance Requirements for Supplies and Consumables

Suggested Content:
Description of how and by whom supplies and consumables will be inspected and accepted

<u>What is the purpose of this element</u>? The purpose of this element is to establish and document a system for inspecting and accepting all supplies and consumables that may directly or indirectly affect the quality of the project or task. If these requirements have been included under another section, it is sufficient to provide a reference.

<u>What type of information should be included in this element</u>? Geospatial projects may require the use of supplies and consumables such as film, photography paper, or batteries that need to be checked to assure they meet requirements. Clearly identify such supplies or consumables to be used on the project. Document the acceptance criteria by which the supplies or consumables will be judged, the procedures used to test the materials and consumables, and the frequency of these tests. Finally, document the corrective actions to be taken in case supplies or consumables do not meet acceptance criteria.

If a geospatial component of a larger environmental sampling project exists, consumables and supplies used during sample collection would be included in the QA Project Plan for the environmental sampling portion of the project.

3.2.9 B9. Data Acquisition Requirements (Nondirect Measurements)

> **Suggested Content**:
> - Description of secondary data used
> - Description of the intended use of the data
> - Acceptance criteria for using the data in the project and any limitations on that use

<u>What is the purpose of this element</u>? Quality assurance includes not only the collection of new data, but also an evaluation of any existing data used. The secondary use of existing data (or "nondirect measurements") is an important component of many geospatial data projects. These data are to be evaluated to determine that they are of adequate quality for the project's needs. This element documents the sources of data and the criteria used to evaluate the quality of this data.

<u>How is "secondary use of existing data" defined and what are some examples for geospatial data projects</u>? Almost every geospatial project makes use of existing data, because data collection is resource intensive and time consuming. Collecting new geospatial data can be avoided by using existing sources of geospatial data developed by local, state, and federal agencies, as well as commercial data vendors. The most common types of commercially available geospatial data are up-to-date street centerline files (with accurate address ranges) and satellite imagery from commercial vendors. Various federal agencies generate and supply large quantities of geospatial data that are used throughout the country; examples include Digital Line Graphs, Digital Elevation Models, the National Land Cover Database, and the National Hydrography Dataset.

<u>What is the purpose of the acceptance criteria for secondary use of existing data, and what are some specific criteria to consider</u>? Criteria would be developed to assure existing data from other sources is of the type, quantity, and quality needed to meet the project's product objectives. These criteria would be documented in the Data Acquisition Requirements (Nondirect Measurements) (B9) element. Examples of these criteria include:

- project-specific requirements for content and accuracy of data to be acquired;
- standards for metadata needed for the planned data quality assessments;
- acceptable coordinate systems:
  - projection
  - units
  - datum
  - spheroid;
- acceptable data formats (One way of documenting this is to indicate that any format supported as a transfer format by the GIS software system is acceptable, particularly if the best source of data for the project is from a computer-aided design package,

because extensive editing and manipulation could be required to convert the data into an acceptable format.);
- acceptability criteria of non-GIS sources (zip code lists, latitude/longitude lists) from spreadsheets or database files;
- acceptable levels of data loss if any data conversion is to be done;
- the geographic coverage requirements (e.g., Does the external data to be assessed cover the study area? This is especially relevant for projects with study areas in AK, HI, Guam, or other U.S. Territories.);
- how limitations of these data are to be documented.

Additional items to consider when writing this element include the following:

- To the extent that they are known, "gray" areas in the use of the data in the project would be documented here. For example, if the only available data source is at a scale or accuracy that is questionable for its intended use, make sure these concerns/ limitations are documented and the potential effects on the final data are known. If this analysis has not yet been completed when the QA Project Plan is being developed, this element would contain directions for documenting this information.

- If an outside service (such as commercially available geo-coding companies) is to be used to produce geographic coordinates from addresses, define the acceptable limits for completeness and accuracy of matching and document their data processing procedures.

For remote-sensing data sets, similar criteria and assessments would be provided in this element. In addition, the level of processing (and the product) would be identified and documented in the task for the commercial vendor.

The Data Acquisition Requirements (Nondirect Measurements) (B9) element of the QA Project Plan would clearly identify the intended sources of previously collected geospatial data or imagery to be used in the project. The care and skepticism applied to the generation of new data are also appropriate to the use of previously compiled data. For example, EPA risk assessment and risk management analyses use spatial interrelationships of natural resources, human populations, and pollution sources by processing existing geospatial data within GIS. If data are inappropriate due to scale, accuracy, resolution, or content, this may lead to inappropriate products and decision errors. The quality of the outputs is dependent upon the quality of the input data, as well as the project's data management and processing software/hardware configuration, including documentation and metadata.

The Data Acquisition Requirements (Nondirect Measurements) (B9) element would also include a discussion of limitations on the use of the data and the nature of the uncertainty of the data.

For many of the most commonly used geospatial data (such as U.S. Geological Survey Digital Line Graph layers, Digital Elevation Model data, or National Land Cover data), the existing metadata are the end user's only source of information about the accuracy, content, usefulness, and completeness of the data. The user will evaluate these existing data sources against the requirements of the project using the supplied metadata. Evaluation criteria are set to determine the minimum acceptable quality of data that can be used. The Data Acquisition Requirements (Nondirect Measurements) (B9) element would contain instructions for documenting any effects of compromises made in order to use the data.

How should quality issues be documented when using, combining, or analyzing data from different sources? This element of the QA Project Plan would contain guidance on combining different data sources from widely different scales. For example, if the project is to identify the parcels in a city that are within a floodplain boundary, two types of data might be used: geospatial parcel data and floodplain boundaries. Geospatial parcel data are usually of very high accuracy and precision because they represent legal property boundaries. Floodplain boundaries are frequently less accurate by their very nature. A floodplain boundary is usually defined as the point to which the water will rise given a rainfall episode that is likely to occur once in 50 years or once in 100 years. The floodplain boundary does not represent any actual physical or environmental boundary—it only represents the probable location of a boundary based on statistical analysis of historical rainfall data. The uncertainty resulting from combining these data sets would be documented so that users of the resulting analysis (geographic overlay of parcels and flood zones) will understand how to evaluate any decisions made.

How is metadata used in quality assurance? As mentioned above, metadata are virtually the only source of information about the quality and accuracy of existing data. Candidate geospatial data sets may not have metadata if they were created prior to the development of the 1995 FGDC standards. External data sources may need to be contacted to determine data availability, condition, and constraints on their use. If only partial documentation is obtained, the risk to project objectives of using data of unknown quality would need to be considered.

What other issues might be described in this element? The Data Acquisition Requirements (Nondirect Measurements) (B9) element could also be used to document and evaluate the ability of the hardware/software configuration to handle existing data sources chosen for use in the project. The data structure, media storage form, and platform requirements can be critical to data processing and, therefore, the analyses to be performed in the project. For example, some older data sets were created using formats that are not easily transformed into those useable by the Agency's standard spatial analysis software. It is also important to consider whether the acquired data are current and what the prospects are for continued updating to assure future usefulness.

Logical consistency of acquired data is particularly important because it can affect data processing and project results. Logic is based on thematic correlations providing the basis for internal validity of a spatial data set, the types of errors encountered can usually be characterized as systematic (i.e., bias), random, or a simple blunder (Veregin, 1992). Incompleteness of attribute data and loss of data integrity can result in inconsistency of the relationships among the encoded features.

Logical consistency of multivariate data sets of environmental attributes can be screened by statistical tests to evaluate characteristics such as the amount and distribution of missing data, statistical parameters (e.g., sample mean, standard deviation, and coefficient of variation), and data distributions; out-of-range values for the measurement scales; and correlations (EPA, 2000b).

Logical consistency checks can be performed within a geospatial database (e.g., ensure that no parcels in a parcel database have a "development status" code of "undeveloped" along with a "number of buildings" attribute greater than 0, because this is logically inconsistent). Logical consistency checks can also be performed between geospatial databases (e.g., given a set of latitude/longitude coordinates of industrial stacks, ensure that none of them are located in a water feature when overlaid into a land use or hydrography layer).

The Data Acquisition Requirements (Nondirect Measurements) (B9) element would be used to document checks performed on the existing data by the data producers, or, in the absence of such information from the data producer, this element can be used to develop descriptions of the most important checks to perform on the data to ensure that they are usable in the project.

How does one assess the accuracy of geospatial data sets—especially vector data sets? For example, what is the accuracy of the U.S. Census Bureau's Topologically Integrated Geographic Encoding and Referencing (TIGER) data? This is a difficult question to answer; it would be answered by reviewing available metadata and processing information and applying professional judgment to assess the accuracy based on this information.

3.2.10 B10. Data Management

What is the purpose of this element? This element presents an overview of the operations, calculations, transformations, or analyses performed on geospatial data or their attributes throughout the project. Diagrams and graphics illustrating the

Suggested Content:

- Description of the project management or activities
- Flow charts of data usage and processing
- Description of how data will be managed to reduce processing errors
- Description of the mechanism for detecting and correcting errors in data processing
- Examples of checklists or forms to be used
- Description of the hardware/software configuration to be used on the project
- Description of the procedures that will be followed to demonstrate acceptability of the process
- Description of the data analysis or statistical techniques to be used

sources of each data set, the steps through which each one will be processed (including combinations to create new data sets), the names and characteristics of interim data sets, and the naming conventions used at each step can be used to illustrate the processing methodology. The Data Management (B10) element would document operations performed on the data at each step of the process (Figure 5).

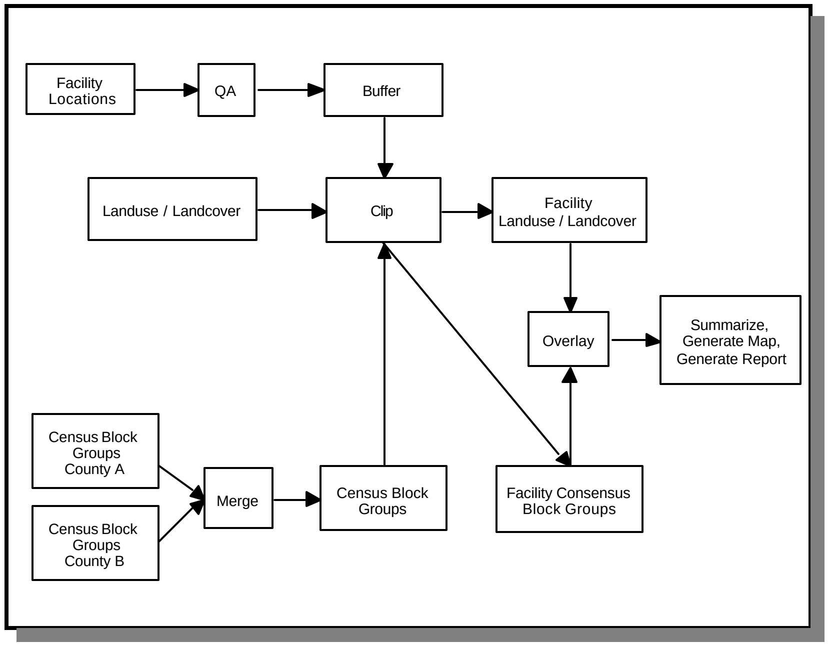

Figure 5. GIS Flow Diagram

<u>What type of information might be included in this element</u>? The Data Management (B10) element includes a discussion and description of records kept throughout the project. This is similar to what would be included in the Documents and Records (A9) element, but includes more detailed descriptions of data set names and processing methods. The Data Management (B10) element might also discuss the requirements for internal program documentation (that is, programmers' comments included with programs). Describe how analysts and others such as software developers will document their work and the steps they take during the course of the project to acquire, analyze, and manage the geospatial data or develop needed software. Describe the function of these notes at the end of the project. For example, when final reports are created to document the overall project and its conclusions, processing notes created by the analysts and managers can provide the actual data processing steps, preserving them to the level of detail required to fully understand the project's technical details or to recreate the product.

The documentation in the Data Management (B10) element might start by describing the process of data management for newly collected geospatial data sets that will undergo data processing in the project. Describe the activities that generate new geospatial data sets through data processing, the use of digitizing tables to render GIS layers from hard-copy map sources, or the synthesis of new data sets from existing data and newly collected data.

What would be covered in this element for geospatial data sets newly collected by GPS?

1. *Define and create data dictionary.* The Data Management (B10) element documents the data dictionary itself. The data dictionary defines the acceptable attributes and codes to be collected during fieldwork. For example, if the project involves collecting information on the location and type of outfall pipes, the data dictionary might include a description of fields used to store pipe material, pipe size, pipe status, and so on. For each of those data fields, coded values would be defined in the data dictionary to restrict the data collector to data using specific, predetermined, valid codes. This would reduce post-processing and cleanup when the data are uploaded to the GIS and would ensure that the correct information is collected in the field.

2. *Transfer the data dictionary to the GPS units.* On many modern GPS units, the electronic data dictionary can be transferred so that the acceptable coding values are accessible in the field. The process by which the data dictionary will be transferred and checked once transferred would be described in the Data Management (B10) element.

3. *Collect and transcribe field notes.* Field notes from data collectors are to be collected and transcribed for use during the data processing and data quality control process. The Data Management (B10) element would document how the notes will be collected, who will collect them, who will input the notes in a form for use by others, and what format and software will be used to store the notes. In addition, the steps and procedures for using the field notes to check data discrepancies and for noting questions during the data transfer and processing steps would be described.

4. *Download the GPS data into the GIS.* Use the Data Management (B10) element to describe the process by which GPS data will be downloaded on the GIS processing computers, and list steps for backing up the raw data and ensuring that it was transferred completely and successfully. The description would also include the procedures for converting the coordinate data into GIS databases, for converting the attribute data into database files, and for reintegrating these data with the coordinate data.

5. *Correct the GPS coordinates* (if necessary). Describe the process to be used to perform the differential corrections on the raw GPS coordinates. If a base station or other GPS unit was used to collect the appropriate reference information, describe the details of the process. Describe any procedures used to check for outliers or other problems created when averaging multiple data locations into a single aggregated location. These types of checks might include calculating the standard deviation of each set of points to be averaged and then checking the standard deviations to make sure none are greater than the specified accuracy criteria.

6. *Document the method, accuracy, and description data for the GPS coordinates.* The method, accuracy, and description data would be integrated into the metadata for the processed, final GPS data sets[1].

What would be covered in the Data Management (B10) element for a map digitizing project?

- Descriptions of how the maps will be prepared for digitizing (e.g., Will Mylar overlays be used to extract the appropriate linework from maps? If so, what will the procedure be?).

- A description of which lines or other information will be extracted from the maps.

- The procedure for assigning identifiers to the features to be digitized.

- A description of the georeferencing identifiers (tics) that might be used to transform the digitized data into geographic coordinate systems.

- Procedures to check the completeness and accuracy of the digitizing effort (Section 3.3).

- The tolerances to be used on the digitizing transformations. For example, when re-registering maps to a digitizing table, what is the acceptable root mean square value to determine whether or not the registration was accurate enough? The root mean square value would also be indicated in the Quality Objectives and Criteria (A7) element as a quality criteria.

By documenting and specifying these types of procedures and tolerances, the digitizing process will go more smoothly and will result in data that require less correction and editing. Similar descriptions explaining how the nonspatial data (attributes) will be collected from the maps, entered into a database, and linked up with the spatial data would be included in the Data Management (B10) element.

Group C elements (Section 3.3) would be used to describe how these data (both spatial and nonspatial) are to be checked and corrected. The Data Management (B10) element would be used to document processing and data management methodologies.

[1]Note that the EPA Locational Data Policy is being reviewed in light of the FGDC metadata guidelines and Executive Order 12906. As the EPA Locational Data Policy is updated, the Latitude/Longitude Data Standard may also be revised to add enough new codes to achieve minimum compliance with FGDC guidelines. See *oaspub.epa.gov/edr/EPASTD$.STARTUP* for status. Extramural organizations (non-EPA), may need to request this document from their EPA work assignment manager.

When existing data (acquired from an external source) are to be used on the project, what might be included in the Data Management (B10) element to describe how these data will be managed during the course of the project?

- The procedures to be used to back up the raw data.

- The procedures to be used to construct the GIS database from these data sources (for example, if multiple geographic data sets are required to cover the study area, describe how each data set will be projected and/or transformed into a common coordinate system, how the data sets will be appended together to create a single seamless layer, and what will be done with the resulting layers during the course of the project.).

- Descriptions of how quality of these processes will be assessed and problems corrected will be addressed in Group C elements.

- The procedures to be used to process and analyze these data (for example, detailed flow charts indicating the procedures to be used at each step of the process and explicitly defining the input and output data for each step).

- Definitions of naming conventions for geospatial data sets—during the course of the project many interim data products may be created; by defining and using a system of naming conventions, data management is improved.

What would be included in the Data Management (B10) element to discuss the development and creation of project-specific applications programs or subprograms? For projects involving the development of applications programs that combine underlying GIS commands or operations, document the name, purpose, and functions of each program. Documentation of these programs ("macros") provides additional information about specific operations to be performed during the project. Many of these procedural programs are developed during the course of the project—not before. The Data Management (B10) element creates a placeholder for descriptions of these macros. Because macros are a prime operational tool in geospatial projects, they are to be developed, documented, and checked carefully. Many of the quality errors that crop up unexpectedly at the end of geospatial projects are due to errors in macro programs that are not caught and corrected early in the process. Use the Group C elements (Section 3.3) to describe how macro programs will be evaluated to ensure that they produce results of the quality indicated in the Quality Objectives and Criteria (A7) element.

The Data Management (B10) element provides guidance to GIS analysts and technicians for properly testing informal macro programs. The Data Management (B10) element could also be used to describe the process whereby macro programs will be checked by senior analysts or QA Officers to ensure that they are working correctly. The Data Management (B10) element might also be used to specify where data are stored and managed on computers, including path names to project files.

Security

Security is an important aspect of data management and quality assurance in general, because security problems may affect the data quality and data usability. The Data Management (B10) element may be used to describe procedures and issues related to the following:

- ***Internet Security:*** Internet security is an important issue in geospatial projects that use the Internet to acquire or transmit data. Describe potential problems with acquiring or transmitting data caused by Internet firewalls. For example, if acquiring existing data from EPA, will access to data within EPA's firewall be a problem?

- ***Confidential Business Information:*** Highly detailed and legally binding procedures are required when working with data designated as Confidential Business Information. If geospatial data (or related attribute data) have been labeled as Confidential Business Information, the appropriate procedures are to be followed. In addition, the Data Management (B10) element could be used to document and describe how the application of Confidential Business Information procedures will affect data access, and therefore, the project timeline.

- ***General Computer and Physical Plant Security:*** The Data Management (B10) element could be used to describe any special considerations, procedures, or characteristics of the computing environment or physical plant that might affect the security of the data being processed on the project. For example, if there are special considerations regarding user access rights to particularly sensitive data, the Data Management (B10) element could be used to document these issues.

Electronic Exchange Formats

When the results of a geospatial project are to be transmitted to other data users in the organization or to external organizations, the Data Management (B10) element would be used to document the formats to be used for the data exchange.

Hardware/Software Configuration

What might be the general structure of the discussion of the hardware/software configuration presented in this portion of the QA Project Plan? The discussion of hardware/software configurations will depend on the purpose of the subprograms to be developed on the project. If the purpose of the overall project is to develop GIS or geospatial software for a wider audience of users beyond the project team itself, then it would be helpful for the QA Project Plan to take into account EPA policies regarding software development, life-cycle planning, and other policies outlined in the *Information Resources Management Policy Manual* (EPA, 1998b).

For projects where applications programs or processing programs are developed solely for use as data processing enablers on the project, the Data Management (B10) element may be used to describe the hardware and software configuration under which the project will be performed. For example, discuss the computer hardware configuration for the project and discuss GIS or other geospatial software required to perform the data processing.

What might be included in the QA Project Plan for geospatial software development projects whose purpose is to develop a standardized software product for an audience beyond the project team? For these projects, the Data Management (B10) element would be used to discuss the major design issues of the software. However, the Data Management (B10) element would supplement, not replace, a formal software design and development methodology in which the details of the software's design and operation would be documented.

This element may also address performance requirements (e.g., run times) and other features that characterize or assess the hardware/software configuration. This discussion could be incorporated within a general overview of the configuration's QA program. [Assessments that target the GIS software itself and its ability to process geospatial data are addressed by the Group C elements within the QA Project Plan (Section 3.3).] The configuration's QA program is jointly planned and implemented by the project management team and the software developer's independent QA staff, generally as part of systematic planning [the Quality Objectives and Criteria (A7) element]. It addresses the use of standards, test planning and scheduling, level of documentation required, personnel assignments, and change control. It also ensures that timely corrective action will be taken as necessary. Items within the systems development life cycle that are relevant to the particular modeling project may also be considered when establishing the configuration's QA program. Examples of such items, taken from Chapter 4 of EPA's *Information Resources Management Policy Manual*

(Directive 2100) (EPA, 1998b) and the Information Technology Architecture Roadmap,[2] are provided in Table 3.

What important issues would the QA Project Plan address for the hardware/software configuration's QA program? It is important that the QA Project Plan specify the particular QA procedures that will be implemented within the software development project to ensure that the data generated by the product are defensible and appropriate for the planned final use. This section of the QA Project Plan would address QA efforts performed as the data management and processing systems are being developed. These efforts may include:

- identifying necessary requirements for the hardware/software configuration and establishing quality criteria that address these requirements within the systematic planning and needs analysis phase of the project [Quality Objectives and Criteria (A7) element];

- implementing an appropriate project management framework to ensure that the requirements and quality criteria established for the hardware/software configuration are achieved [as discussed in the Project Management Group (A4-A9) elements and the Data Acquisition Requirements (Nondirect Measurements) (B9) element];

- performing testing and other assessment procedures on the configuration to verify that the requirements and quality criteria are being met [details on the assessment procedures are addressed in the Assessment Methods and Response Actions (C1) element].

The magnitude of these QA efforts will depend on the underlying complexity of the geospatial data effort and the required hardware/software configuration. Therefore, EPA's graded approach (Chapter 1) will direct the overall scope of these QA efforts.

---

[2]Published by EPA's Office of Technology Operations and Planning, formerly the Office of Information Resources Management, Directive 2100 establishes a policy framework for managing information within EPA. It can be accessed online at *www.epa.gov/irmpoli8/polman/index.html.* The Information Technology Architecture Roadmap, which contains annual updates of this document, can be found at (internal EPA web site) *Basin.rtpnc.epa.gov:9876/etsd/ITARoadMap.nsf.*

Table 3. Typical Activities and Documentation Prepared Within the System Development Life Cycle of a Geospatial Data Project to Be Considered When Establishing the QA Program for the Hardware/Software Configuration

Life Cycle Stage	Typical Activities	Documentation
Needs Assessment and General Requirements Definition	• Assessment of needs and requirements interactions in systematic planning with users and other experts	• Needs assessment documentation (e.g., in the QA Project Plan, if applicable) • Requirements document
Detailed Require-ments Analysis	• Listing of all inputs, outputs, actions, computations, etc. that the geographic information or modeling system is to perform • Listing of ancillary needs such as security and user interface requirements • Design team meetings	• Detailed requirements document, including performance, security, user interface requirements, etc. • System development standards
Framework Design	• Translation of requirements into a design to be implemented	• Design document(s), including technical framework design, software design (algorithms, etc.)
Implementation Controls	• Coding and configuration control • Design/implementation team meetings	• In-line comments • Change control documentation
Testing, Verification, and Evaluation	• Verification that the software code, including algorithms and supporting information system, meets requirements • Verification that the design has been correctly implemented • Beta testing (users outside QA team) • Acceptance testing (for final acceptance of a contracted product) • Implement necessary corrective actions	• Test plan • Test result documentation • Corrective action documentation • Beta test comments • Acceptance test results
Installation and Training	• Installation of data management system and training of users	• Installation documentation • User's guide

Table 3. Typical Activities and Documentation Prepared Within the System Development Life Cycle of a Geospatial Data Project to Be Considered When Establishing the QA Program for the Hardware/Software Configuration

Life Cycle Stage	Typical Activities	Documentation
Operations, Maintenance, and User Support	• Usage instructions and maintenance resources for geographic information or model system and databases	• User's guide • Maintenance manual or programmer's manual
System Retirement and Archival	• Information on how data or software can be retrieved if needed	• Project files • Final report

<u>How are requirements and criteria placed on the hardware/software configuration addressed in systematic planning</u>? Elaborating further on the first bullet above, the systematic planning phase of the study [Quality Objectives and Criteria (A7) element] defines requirements and quality criteria for the data processing system to ensure that the project's end-use needs can be adequately met. For example, criteria on errors propagated by data processing would be established during systematic planning to ensure that uncertainty requirements for the mode outputs can be met. Such requirements and criteria, therefore, impact the project's hardware/software configuration.

In systematic planning, questions such as the following may be addressed when defining these requirements and quality criteria:

- What are the required levels of accuracy and uncertainty for numerical approximations?
- Are the selected mathematical features of the program (e.g., algorithms, equations, statistical processes) appropriate for the program's end use?
- Are the correct data elements being used in the calculations performed within the program's algorithms?
- What requirements regarding documentation and traceability are necessary for the program's inputs, interim outputs, and final outputs?

Other items addressed during systematic planning that are likely to impact assessment of the hardware/software configuration include security, communication, software installation, and system performance (e.g., response time). These issues are addressed briefly below.

<u>What kinds of documentation might the QA Project Plan address as part of hardware/software configuration for a software development project</u>? When documenting planning and performance components of hardware/software configuration, project and QA Managers may tailor the documentation to meet the specific needs of their project. Examples of different types of

documentation that can be generated for various tasks within the planning phase of the system's life cycle include the following:

- *Requirements Documentation* (IEEE, 1998): The general requirements document gives an overview of the functions that the model framework will perform.

- *Design Documentation*: Design documents plan and describe the structure of the computer program. These are particularly important in multiprogrammer projects in which modules written by different individuals interact. Even in small or single-programmer projects, a formal design document can be useful for communication and for later reference.

- *Coding Standards or Standard Operating Procedures*: These may apply to a single project or a cumulative model framework and need to be consistent across the development team.

- *Testing Plans (FIPS 132[3])*: Testing is to be planned in advance and is to address all requirements and performance goals.

- *Data Dictionary*: A data dictionary can be useful to developers, users, and maintenance programmers who may need to modify the programs later. The data dictionary is often developed before code is written as part of the design process.

- *User's Manual*: The user's manual can often borrow heavily from the requirements document, because all the software's functions would be specified there. The scope of the user's manual would take into account such issues as the level and sophistication of the intended user and the complexity of the interface. Online help can also be used to serve this function.

- *Maintenance Manual*: The maintenance manual's purpose is to explain a framework's software logic and organization for the maintenance programmer.

- *Source Code*: It is very important to store downloadable code securely and to archive computer-readable copies of source code according to the policies of the relevant regulatory program.

---

[3] Federal Information Processing Standards

- *Configuration Management Plan* (IEEE, 1998): The configuration management plan provides procedures to control software/hardware configuration during development of the original software and subsequent revisions.

Additional information and examples can be found in Chapter 17 of EPA's *Information Resources Management Policy Manual* (Directive 2100) (EPA, 1998b). In general, it is best to coordinate any discussion of documentation in the QA Project Plan with information presented in the Documentation and Records (A9) element.

What kinds of standards do I include in the hardware/software configuration's QA program to ensure that the configuration is compliant and acceptable? The configuration is to be designed to comply with applicable EPA information resource management policies and data standards, which can be found within EPA's *Information Resources Management Policy Manual* (Directive 2100) (EPA, 1998b). Other standards may also be applicable and are to be cited, such as the Federal Information Processing Standards, which govern the acquisition of U.S. Government information processing systems. This element of the QA Project Plan is the place to introduce these standards and discuss how the project will ensure that they will be addressed.

Sources for determining specific types of standards include the following:

- EPA's *Information Resources Management Policy Manual* (Directive 2100) (EPA, 1998b) includes EPA hardware and software standards to promote consistency in use of standard support tools such as computer-aided software engineering tools and coding languages, as applicable, by contractors and EPA staff in GIS software development and maintenance efforts.

- Chapter 5 of EPA's *Information Resources Management Policy Manual* (Directive 2100) (EPA, 1998b) defines applicable EPA data standards.

- EPA's Environmental Data Registry (*www.epa.gov/edr*) promotes data standardization, which allows for greater ease of information sharing.

- The EPA Information Technology Architecture Roadmap provides guidance for the selection and deployment of computing platforms, networks, systems software, and related products that interconnect computing platforms and make them operate.

- Publications on Federal Information Processing Standards govern the acquisition of U.S. Government information processing systems.

Directives and standards such as these are frequently revised. Therefore, it is important that these directives and standards be reviewed frequently to ensure that the latest versions are being utilized. See *oaspub.epa.gov/edr/EPASTD$.STARTUP* for standard status. Extramural organizations may check with their EPA work assignment manager for current status. The QA Project Plan is to specify how the configuration will be verified or demonstrated according to these and other standards.

3.3 GROUP C: ASSESSMENT/OVERSIGHT

Group C elements are used to document the process of evaluating and validating the data collection and data processing activities on the project. In other words, Group C includes descriptions of the quality assessments and evaluations, and describes the reports and actions to be taken, based on assessments.

Whereas Group B elements describe the methods of collecting geospatial data types and methods of choosing and managing geospatial data sources, Group C elements focus on the quality assessments that will be performed during the data processing of the project. In addition, Group C is used to describe the procedure for addressing quality problems.

There is some overlap between discussions in the Data Management (B10) element and those in Group C. This is because data management and the programs used to manage and process geospatial data are the root of many of the quality problems. However, Group C is to be used to augment the Data Management (B10) element when using existing data and to describe the steps taken to ensure that assessments in the Data Management (B10) element and other parts of the QA Project Plan are implemented.

3.3.1 C1. Assessments and Response Actions

<u>What is the purpose of this element</u>? This element describes the internal and external checks necessary to ensure that:

- all elements of the QA Project Plan are correctly implemented as prescribed;
- the quality of the data and product generated by implementation of the QA Project Plan is adequate;

Suggested Content:

- Description of each assessment
- Information expected and success criteria
- Assessments to be done within the project team and which are done outside the project team
- The scope of authority of assessors
- Discussion of how response actions to assessment findings are to be addressed
- Description of how corrective actions will be carried out

- corrective actions, when needed, are implemented in a timely manner and their effectiveness is confirmed.

What type of information might be included in this element? Based on the project's quality needs, scope, and limitations on uncertainty, different levels of assessments and response actions may be appropriate. For each of the assessments described in the Assessment and Response Actions (C1) element, include a description of activities that will be used to correct problems or errors, as applicable.

The following types of assessments would be documented in the Assessment and Response Actions (C1) element as a means of ensuring that secondary data being evaluated meet the specifications noted in the Quality Objectives and Criteria (A7) and Data Acquisition Requirements (Nondirect Measurements) (B9) elements:

- Check locations of features in existing data against locations of these features in other data sources. For example, describe how digital elevation model elevations will be spot-checked against topographical maps, to ensure that the accuracy of the digital elevation models is within its accuracy specifications.

- Check attribute data to ensure that it is of acceptable quality, based on the criteria specified in the Quality Objectives and Criteria (A7) element (see Appendix C for more information).

- Describe how senior level scientist/GIS analysts will review processing procedures during methodology development. Identify potential processing problems, issues, and work-arounds.

- Describe the requirements for reviewing data at the end of each processing step. Are data consistent? Are data values correct given the processing manipulation performed? Are the locations of geographic entities within expected norms based on processing techniques employed? If macros or other data processing programs are run, describe how data inputs and outputs will be tested to ensure that their characteristics are as expected and that the programs performed the functions defined for them.

- Describe the methods used to compare, evaluate, and assess the data produced in each step of the project to ensure that they have been processed correctly. When macros are used to automate a multistep process, code the macro in such a way that the results of each step can be independently examined so that, if problems are found in the final output data set, the error can be found by reviewing data at each prior step in the process.

- Use the Assessment and Response Actions (C1) element to describe tests that compare processed geospatial data to the original or source data sets throughout production. Describe expected changes in the data and unexpected or erroneous changes. For example, when converting from raster to vector data formats, compare the vectorized data to the original raster data to ensure that the appropriate cell size was used and that no transformations or inappropriate aggregations occurred. When converting from vector to raster, describe how the raster data set's cells would be coded when original vector lines divide the raster cells. Will the vector polygon having the greatest area be used for the cell code, or will the cell be coded using an average of the values in the coincident polygons?

- Describe how the assessments will ensure that no geographic features or data were lost, deleted, or removed unexpectedly. Loss of geographic features can be an issue when tolerances are inappropriately applied, resulting in coalescence of geographic features. Identify methods of ensuring that the right number of features are present at each step of the process; by doing so, problems with feature loss due to inappropriate tolerances can be determined.

- Even in projects having limited scope or complexity, it may be appropriate to describe the procedures used to design, develop, and test macro programs during the course of the project. Use the Assessment and Response Actions (C1) element to document that procedure, especially in light of how the programs will be assessed for proper operation.

- For all assessments, identify who will conduct the assessment, indicating their position within the project's organization.

- Describe how and to whom the assessment information will be reported

.

- Define the scope of authority of the assessors, including stop-work orders and when assessors are authorized to act.

The following is a description of various types of assessment activities available to managers of geospatial projects for evaluating the effectiveness of project implementation.

A. *Readiness review* is a technical check to determine if all components of the project are in place so that work can commence on a specific phase.

These reviews can help avoid redoing expensive field work by assuring that equipment is in proper working order (e.g., charged battery pack, adequate performance of GPS receiver units) and that adequate logistical preparations, such as acquiring supporting materials and property access are performed before a survey.

B. *Technical Systems Audit* is a thorough and systematic, on-site, qualitative audit in which facilities, equipment, personnel, training, procedures, and record keeping are examined for conformance to the QA Project Plan. The technical systems audit is a powerful audit tool with broad coverage that may reveal weaknesses in the management structure, policy, practices, or procedures. It is ideally conducted after work has commenced (such as during image acquisition) but before it has progressed very far. The technical systems audit provides opportunity for corrective action.

For example, technical systems audits are conducted for remote sensing operations by the QA staff of an EPA contractor, or by the Agency itself, to compare observed operations with a set of approved standard operating procedures and QA protocols defined in the QA Project Plan for the work assignment. These audits are facilitated by use of an audit questionnaire designed to systematically guide the auditor through various remote-sensing processes. The questionnaire ensures that all pertinent operations are thoroughly evaluated during the audit. Findings are recorded on a project-specific checklist. Audit reports document appropriateness of operations, note problems and obstacles, and recommend corrective actions to the project manager, who notifies EPA management via a memorandum.

C. *Performance Evaluation* is a type of audit in which the quantitative data generated by a measurement system such as GPS are obtained independently and compared with routinely obtained data to evaluate the proficiency of the sample collector. The QA Project Plan lists the performance evaluations that are planned, identifying:

- the sample to be taken
- the target location to be covered
- the timing/schedule sample duplication
- the aspect to be assessed (e.g., precision, bias).

On a project where new aerial photography is being acquired, for example, the project lead, upon receipt from the photo laboratory, would screen the original film (or contact prints, and/or enlargements) for such parameters as exposure, length of the leader/trailer, and appropriate camera mounting; verify the acceptability of overflight products [i.e., scale (correctness), coverage (completeness), resolution (detection limit)] for photo analysis requirements; and document findings to ensure overall image acceptability.

D. *Surveillance* is the continual or frequent monitoring of the status of a project and the analysis of records to ensure that specified requirements are being fulfilled. It can occur at various steps in the project and be a self-assessment or an independent assessment.

For example, the production of output from the photo laboratory (and/or digital scanning) subcontractor would be monitored to ensure they are able to meet the deliverable date and provide photos enlarged to common scale. Under an umbrella QA Project Plan covering many routine tasks, processes and products could be inspected internally using standardized QA checklists (e.g., film and photography screening photo analysis reports) documented in monthly reports assessing the progress, performance, and quality of activities.

E. *Audit of Data Quality* reveals how the data were handled, what judgments were made, and whether uncorrected mistakes were made. Performed prior to producing a project's final report, audits of data quality can often identify the means to correct systematic data reduction errors.

For example (or at the minimum), a formalized procedure would be described for quality assessment during implementation of a project processing geospatial data (whether collected or acquired) on a GIS to prepare a product. Describe assessment and response activities to ensure the quality of the product, including review of the acquired data or images assessment reports [Data Acquisition Requirements (Nondirect Measurements) (B9) element] to ensure that the lineage is traceable and defensible for the type of information required. If inadequacies are identified, the data analyst would contact the project's data producer to correct any identified problems, or if the data were acquired from an outside source, a different data set may need to be acquired for processing. Any problems identified and corrective actions taken would be documented to ensure that the project requirements are satisfied. Reviews of the interim steps in data reduction or transformations by an independent analyst are also needed prior to the product's completion to confirm adequacy of reductions and transformations and to confirm that topology is established properly for the data set. Any problems identified in the data set produced by the project or omissions in documentation identified by these reviews need to be corrected before the product is completed.

F. *Peer review* is primarily an external scientific review. Reviewers are chosen who have technical expertise comparable to the project's performers but who are independent of the project. Peer reviews ensure that the project activities:

- were technically adequate

- were competently performed
- were properly documented
- satisfied established technical requirements
- satisfied established quality assurance requirements.

In addition, peer reviews assess the assumptions, calculations, extrapolations, alternative interpretations, methods, acceptance criteria, and conclusions documented in the project's report. The names, titles, and positions of the peer reviewers, if known, are to be included in the QA Project Plan and their planned findings report(s). Responsibilities for reports documenting responses to peer-review comments and completed corrective actions would be specified.

For example, project team members review photo interpretations made by the project analyst and the technical supervisor in order to assess and validate the reasonableness and soundness of interpretations.

G. *Data Quality Assessment* involves the application of statistical tools to determine whether the data meet the assumptions under which the data quality objectives and data collection design were developed and whether the total error in the data is tolerable. *Guidance for Data Quality Assessment: Practical Methods for Data Analysis (QA/G-9)* (EPA, 2000b) provides guidance for planning, implementing, and evaluating data quality assessments.

For example, a geospatial data set could be reviewed by an independent analyst to check data quality (e.g., univariate descriptive statistics and outlier tests), logical consistency (e.g., thematic correlations) for internal validity of multivariate data sets, proper topology, and traceable and defensible lineage.

<u>How might the assessments be documented</u>? The number, frequency, and types of assessments would be included in this element. Depending on the nature of the project, there may be more than one assessment. The QA Project Plan would specify the individuals, or at least the specific organizational units, who will perform the assessments. Independent assessments are performed by personnel from organizations not connected with the project but who are technically qualified and who understand the QA requirements of the project.

Audits, peer reviews, and other assessments often reveal findings of practice or procedure that do not conform to the written QA Project Plan. Because these issues need to be addressed in a timely manner, the protocol for resolving them is outlined in this element together with proposed corrective actions to ensure that such actions are performed effectively. The person to whom the concerns are to be addressed, the decision-making hierarchy, the schedule and format for oral and written reports, and the responsibility for corrective action are all discussed in this element. This element also explicitly defines the unsatisfactory conditions upon which the assessors are authorized to act and list the project personnel who are to receive assessment reports.

3.3.2 C2. Reports to Management

Suggested Content:
- Frequency and distribution of reports issued to management that document assessments, problems, and progress
- Individuals or organizations responsible for preparing the reports and actions recipients would take upon receipt of the

What is the purpose of this element? This element provides a place to document the frequency, type, distribution, and content of reports that will record the status of the project and, specifically, data assessments made in the Assessment and Response Actions (C1) element.

What type of information might be included in this element? The graded approach to QA Project Plans implies that, for projects of very limited scope, quality requirements, or size, a simple description of the use of weekly or monthly status e-mails may be appropriate. For more complex projects with many processing steps, data sources, and complex processing methods, more formal reports may be required and documented in the Reports to Management (C2) element.

Effective communication among all personnel is an integral part of a quality system. Planned reports provide a structure for apprising management of the project schedule, deviations from approved QA and test plans, the impact of these deviations on data quality, and potential uncertainties in decisions based on the data. Verbal communication regarding deviations from QA plans would be noted in summary form in the Data Review, Verification, and Validation (D1) element.

No matter how informal or formal the reports may be, it is appropriate to describe the content, frequency, and distribution of these reports in the Reports to Management (C2) element. This element would also identify the individual or organization responsible for preparing the reports and action recommendations that might be included in the reports. An important benefit of the status reports is the opportunity to alert management to data quality problems, propose viable solutions, and procure additional resources. If the project is not assessed continually (including evaluation of the technical systems, measurement of performance, and assessment of data), the integrity of the data generated in the project may not meet quality requirements. Submitted in a timely manner, these assessment reports will provide an opportunity to implement corrective action when most appropriate. At the end of a project, a report documenting the data quality assessment findings is submitted to management.

3.4 GROUP D: DATA VALIDATION AND USABILITY

Group D elements describe final data validation and usability procedures used to ensure that the final product meets quality and completeness criteria. Because geospatial projects involve a great deal of data processing, frequent manipulations of geospatial data, and sometimes extensive software development, many assessments may be carried out during the course of the project. These types of assessments would be documented in the Data Management (B10) element and in the Assessment and Response Actions (C1) element. Group D elements facilitate examination of the final data product or cartographic product to ensure that it is of acceptable quality and can be used for its intended purpose.

The process of data verification requires confirmation by examining or providing objective evidence that the requirements of these specified QC acceptance criteria are met. In design and development, verification concerns the process of examining the result of a given activity to determine conformance to the stated requirements for that activity. The process of data or imagery verification effectively ensures the accuracy of data, using specified methods and protocols, and is often based on comparison with reference or control points and base data.

The process of data validation requires confirmation by examination and provision of objective evidence that the particular requirements for a specific intended use have been fulfilled. Validation, usually performed by someone external to the data generator, is the process of examining a geospatial product or result to determine conformance to user needs.

3.4.1 D1. Data Review, Verification, and Validation

Suggested Content:
- The criteria to be used to validate and verify the final product

What is the purpose of this element? This element would be used to describe the criteria that will be used in accepting or rejecting the final product. Many of these criteria may be gleaned from assessments and checks identified in other portions of the QA Project Plan. However, in the Data Review, Verification, and Validation (D1) element, pay close attention to those criteria that would make the data inappropriate for its intended use. When producing a final product in a geospatial project, many quality checks and assessments are carried out during production [as described in the Data Management (B10) and Assessments and Response Actions (C1) elements], but the final product itself would also undergo final checks to ensure that it meets the objectives for usability and quality.

What type of information might be included in this element? For data collection involving GPS surveys or aerial photography, note how closely the coordinates or imagery represent the actual surface feature and whether or not that difference is within acceptable tolerances. By noting deviations in sufficient detail, subsequent data users will be able to determine the data's usability under scenarios different from those included in project planning. The strength of conclusions that can be drawn from data (see *Guidance Document for Data Quality Assessment: Practical Methods for Data Analysis (QA/G-9)* (EPA, 2000b) has a direct connection to the sampling design and deviations from that design. Where auxiliary variables are included in the overall data collection effort (for example, groundwater or ozone data), they would be included in this evaluation. [Environmental data are covered in *Guidance for Quality Assurance Project Plans (QA/G-5)* (EPA, 1998a).]

How would sample collection and handling procedures or deviations be handled? Details about the acquisition of geospatial samples and imagery are important for properly interpreting the results. The Sampling and Image Acquisition Methods (B2) element provides these details, which include sampling or imagery acquisition procedures and equipment (e.g., camera and film type, control points). Acceptable departures (for example, alternate GPS sampling sites) from the QA Project Plan, and the action to be taken if the requirements cannot be satisfied, are to be specified for each critical aspect. Validation activities would note potentially unacceptable departures from the QA Project Plan. Comments from field surveillance on deviations from written field survey or flight plans would also be noted.

What type of quality control steps would be performed in this element? The Quality Control (B5) element of the QA Project Plan specifies the QC checks that are to be performed during sample collection, handling, and analysis. These include analyses of reference data or control points and calibration standards that provide indications of the quality of data being produced by specified components of the measurement process. For each specified QC check, the procedure, acceptance criteria, and corrective action (and changes) would be specified. Data validation would document the corrective actions that were taken, samples or images affected, and the potential effect of the actions on the validity of the data.

When data or materials are acquired from other sources, verify that the materials are received as originally ordered and that the order is complete. For example, for samples taken by GPS technology, the standard deviation of the field data can be checked during the postprocessing data assessment. For imagery, the contents of each photo data package or digital file can be checked for coverage and quality upon completion receipt. If new photographs were acquired, accuracy of elevations and positions would be checked against targets placed on the ground to mark control points in advance of the aerial survey/photography.

Scientists and contractors performing photogrammetric analysis tasks would be expected to adhere to standards such as the National Map Accuracy Standards and other standard operating procedures for data analysis and product generation (e.g., comparison of index point coordinates from the end of a measurement session with those taken at the beginning to see if the discrepancy exceeds digitizer control limits). Positional accuracy of points and associated area perimeters, as well as the methods used to establish them, would be reported in ground control reports as part of a draft photogrammetry report. The latter would be reviewed in the product accuracy assessment to determine if accuracy met project objectives established for data use. Known but withheld coordinates would be used to evaluate the final compilation by comparison to at least one test point established for each project area and carried through in the photogrammetric process. If no targets were established, three or more discrete imaged features would be used as controls and compared to field-survey ground coordinates or comparable features on existing photographs or maps. The residuals or discrepancies between field-established coordinates and the photogrammetric coordinates at two points can be used to indicate a misidentification, with the residual (discrepancy) at the third point identifying any bad (misidentified) point.

If instruments such as GPS receivers, digitizing tablets, or other measurement equipment are used on the project, document the results of calibration activities in this element. Ensure that the calibrations:

- were performed within an acceptable time prior to generation of data or imagery;
- were performed in the proper sequence;
- included the proper number of calibration points.

When calibration problems are identified, any data or imagery produced between the suspect calibration event and any subsequent recalibration would be flagged to alert data users.

3.4.2 D2. Verification and Validation Methods

Suggested Content:
- Description of validation and verification processes for the final products
- Discussion of issues related to resolving problems detected and identification of individuals or authorities who will determine corrective actions
- Description of how the results of the validation will be documented for the product users
- Definition of differences between validation and verification issues

What is the purpose of this element? This element is the appropriate place to describe *how* the final products will be verified and validated. Whereas the Data Review, Verification, and Validation (D1) element documents what final checks will be performed, this element describes how these checks will be carried out.

As with Data Review, Verification, and Validation (D1) element, a substantial amount of the information relevant to this element may be found in other QA elements throughout the QA Project Plan. This element would include many, if not all, of those procedures. However, because Group D elements (including this element) concentrate on verifying and validating the final products, it addresses ways of modifying or adding to previous assessments to ensure that the final product is acceptable.

What type of information might be included in this element? This type of validation and verification might be necessary, for example, when the final product is a database that will be distributed and used by others. Throughout the production or analysis process, a number of QA checks and assessments are carried out to ensure that procedures are being followed correctly. However, at the very end of the process, a series of final checks are to be implemented to make sure the data will be usable by the intended audience. The amount of data validated is directly related to the project data quality objectives. The percentage of data validated for the specific project, together with its rationale, would be outlined or referenced. The QA Project Plan would have a clear definition of what is implied by "verification" and "validation." The type of checks (and their descriptions) might include:

- verifying that each output data set falls into the correct geographic location and has the specified coordinate system and precision;

- verifying that the files to be delivered are of the specified format [for example, if the project defines that the output format is to be compressed Spatial Data Transfer Standard format, the staff member responsible for the Verification and Validation Methods (D2) element would ensure that the each of the output data sets is indeed in Spatial Data Transfer Standard format.];

- verifying that each data set can be unpackaged, uncompressed, or otherwise configured for use by end-users;

- verifying that all of the required database tables and fields are present.

If a map or cartographic product is to be the final deliverable, the Verification and Validation Methods (D2) element would be used to describe how the content of the map will be checked to ensure that it meets the criteria set out in Groups A and B. For example, do the specified layers exist in the map? Is the title correct? Does the legend reflect each of the data layers in the map? Does the map cover the correct geographic extent? Is the scale of the map correct?

Suggested Content:
- Description of how the products or results will be reconciled with requirements defined by the data user or decision maker
- Description of how reconciliation with user requirements will be documented and how issues will be resolved
- Discussion of limitations on the use of the final data product and how these limitations will be documented

3.4.3 D3. Reconciliation with User Requirements

What is the purpose of this element? The purpose of this element is to outline and specify, if possible, the acceptable methods for evaluating the results obtained from the project. This element includes scientific and statistical evaluations of data to determine if the data are of the right type, quantity, and quality to support their intended use.

In most geospatial projects, an abbreviated form of systematic planning addressing acceptance and/or performance criteria rather than a formal DQO Process will be followed. In environmental sampling projects that have a geospatial component, systematic planning would be completed with respect to the media sampling design and analytical methods; associated locational data also need established acceptance and performance criteria against which they can be evaluated.

Data quality assessment follows data validation and verification. This process determines how well validated data can support their intended use. If an approach other than data quality assessment has been selected (e.g., product review), an outline of the proposed activities would be included. For example, graphics products including draft, interim, and final enlargements; scanned photographs; and associated overlays would be reviewed during the internal and external report review process to ensure they meet established graphics standards. The final site analysis report packages would be assessed for quality of site imagery, photo annotations, accuracy of interpreted photographic features, and quality of the associated descriptive text. The editorial quality and consistency of materials included in the report would be evaluated and documented on a QA review checklist.

Data quality assessments of general-purpose databases produced during the course of the project would be compared to quality criteria as specified in Quality Objectives and Criteria (A7), Data Acquisition Requirements (Nondirect Measurements) (B9), and Data Management (B10) elements. For example, on projects where the goal is a database of *georeferenced* water quality locations, the assessment phase would determine whether the final data met the performance criteria (e.g., for accuracy and completeness). The Reconciliation with User Requirements (D3) element would document this comparison and note any deviations that would affect the final product.

Assigning and communicating roles and responsibilities for product reviews [documented in the Project/Task Organization (A4) element] is important. These reviews would, in turn, be coordinated with external QA reviews performed by EPA personnel at the draft and final stages of the report.

CHAPTER 4

GRADED APPROACH EXAMPLES

This chapter is designed to illustrate the structure and content of a geospatial QA Project Plan, providing an example of the elements discussed in Chapter 3. This chapter is important for two reasons: (1) implementation of a new process is always more understandable with examples, and (2) these examples will provide the reader with some insight into the implementation of the EPA graded approach.

> ***The Graded Approach:*** The scope and complexity of the project drive the scope and complexity of the QA Project Plan.

The "graded" approach to developing QA Project Plans means that QA Project Plan development is commensurate with the scope, magnitude, or importance of the project itself. This means that for geospatial projects that are narrow in scope, that will not result in decisions that have far-reaching impacts, or that are not complex, a simple QA Project Plan would be adequate. For complex, broad-scope projects that might lead to regulatory decisions, a more comprehensive and detailed QA Project Plan may be required. Major factors in determining the level of detail needed in the QA Project Plan include the importance of the data, the cost, and the organizational complexity of the project.

> ***Complex Projects:*** Many complex geospatial projects require the development of sophisticated applications or software programs. EPA Directive 2100 (*www.epa.gov/irm_polman*), *The Information Resources Management Policy Manual* (Chapter 17, System Life Cycle Management), categorizes software development projects based on size and complexity.

Geospatial projects usually have a critical software development component as well as the locational data component. The quality issues surrounding software development are also to be taken into account [see *Information Resources Management Policy Manual* (Directive 2100) for more information].

Two aspects of a geospatial project are important for defining the level of QA effort required: *intended use* of the project output and the *project scope and magnitude*. The *intended use* of the geospatial data determines the potential consequences or impacts that might occur because of quality problems. Table 4 shows examples of project data uses frequently encountered in geospatial projects and the corresponding QA issues to address. It is important to attempt to determine the use of the geospatial data or analysis product in the decision-making process to ensure that the data produced are of sufficient accuracy and are of the appropriate type and content to support the decision for which they were created or gathered. Table 4 lists the example projects in decreasing order of the rigor of quality assurance. Final word on the level and degree of rigor for the acceptable level of quality assurance of a specific project lies with the QA Officer.

Table 4. Continuum of Geospatial Projects with Differing Intended Uses

Purpose of Project	Typical Quality Assurance Issues	Level of QA
Regulatory compliance Litigation Congressional testimony	Legal defensibility of data sources Compliance with laws and regulatory mandates applicable to data gathering Legal defensibility of methodology	
...	...	
Regulatory development Spatial data development (Agency infrastructure development)	Compliance with regulatory guidelines Existing data obtained under suitable QA program Audits and data reviews	
...	...	
Trends monitoring (non-regulatory) Reporting guidelines (e.g., Clean Water Act) “Proof of principle”	Use of accepted data-gathering methods Use of accepted models/analysis techniques Use of standardized geospatial data models Compliance with reporting guidelines	
...	...	
Screening analyses Hypothesis testing Data display	QA planning and documentation as appropriate Use of accepted data sources Peer review of products	

As shown in Table 4, projects with a high potential for being involved in litigation (either causing new litigation or being evaluated in ongoing litigation) will generally require a higher level of effort and quality standards in a corresponding QA Project Plan. More modest levels of defensibility and rigor are required for data used for technology assessment or “proof of principle,” where no litigation or regulatory action are expected. Still lower levels of defensibility would be needed for basic exploratory research requiring extremely fast turn-around or high flexibility and adaptability. In such cases, work may have to be replicated under tighter controls or the results carefully reviewed prior to publication. By analyzing the end-use needs, appropriate QA criteria can be established to guide the program or project.

Other aspects of the QA effort can be established by considering the scope and magnitude of the project. The scope of the geospatial project determines the complexity of the QA Project Plan; more complex applications require more QA effort. The magnitude of the project determines the resources at risk if quality problems lead to rework and delays. Data processing projects with nation-wide scope that will produce new Agency-wide data sources (for example, development of the National Hydrography Dataset) would call for sophisticated quality assurance and quality control procedures and extensive QA planning and implementation (and documentation to support evaluation in the secondary use of existing data). Other projects may involve simply acquiring existing digital, geospatial data to create a map in support of management meetings or internal communications. Projects with different scopes are likely to require different levels of QA planning. The level of detail for any particular project is decided by the project's EPA QA Officer. In the case of extramural research, the project's QA Officer will discuss the QA category with the EPA QA Officer so there are no misunderstandings, and any questions will ideally be resolved before work on the QA Project Plan begins. Specific examples of how the considerations described above can be used to define the scope of a project's QA effort are provided below.

In each of the following examples, the information provided under each relevant QA Project Plan element is described to illustrate the application of the element to that example. These examples also discuss the documentation appropriate for each project.

4.1 MINIMUM DOCUMENTATION EXAMPLE: CREATING A CARTOGRAPHIC PRODUCT FROM A SPREADSHEET CONTAINING FACILITY LATITUDE/LONGITUDE COORDINATES

In this example, the geospatial professional has been asked to generate a nation-wide map displaying the locations of certain kinds of industrial facilities based on the locations provided by the requestor in an Excel spreadsheet. Only a subset of the facilities located in the spreadsheet will be mapped. The locations are provided in latitude/longitude format. The subset is identified by a specific code located in a column in the spreadsheet.

4.1.1 Group A: Project Management

Project/Task Organization (A4)—Element A4 would simply state the name, role, and contact information for the geoprocessor performing the work, the person responsible for checking project quality, and the requestor.

Problem Definition/Background (A5)—The geoprocessing professional may have to seek more information from the requestor in order to complete this element. The critical types of information for a limited scope project like this would be as follows:

- Identify the audience for the map.
- Identify and describe the purpose of the map.
- Describe the documentation needed to accompany the map, if any. For example, if the data sources used on the map or the purpose of the map require explanation, document this project-specific requirement.
- Describe the data requirements for the map, including contextual information (for example, state or county boundaries, hydrography, labels) to be included on the map.
- Document any project-specific requirements regarding product disclosure or sensitivity. Describe whether or not the map or the data shown are in any way confidential.

Project/Task Description (A6)—Describe the steps to be taken to complete the project and define, as much as possible, the product to be generated. Things to consider include the following:

- How will the Excel spreadsheet be converted for use in the GIS?
- How will the data be checked for quality?
- Which records will be displayed (if not all)? What is the criteria for selecting specific records to be used in the map?
- What will the map to be generated look like? Include the size, format, title, legend, scale, use of color, and other data to be included (e.g., state boundaries, county boundaries).
- How and when will draft maps be generated, reviewed, and revised?

Quality Objectives and Criteria (A7)—Describe the quality objectives for the project. In a case like this one, example objectives may include the following:

- *The latitude/longitude coordinates in the spreadsheet are to reflect the actual locations of the facilities*. Developing a quality objective like this is important, because the requestor may assume the locations are accurate or precise without having examined them. By including this objective, the geoprocessing professional sets a criterion that can be checked in the assessment phase to address obvious inconsistencies in the latitude/longitude coordinates. For example, some coordinates may only include a latitude/longitude to the closest degree, while others may include latitude/longitude down to a decimal degree. Coordinates that are only precise to a degree of latitude/longitude may be questioned as to their precise representation of an actual facility location.

- *The original latitude/longitude coordinates are to be converted into a GIS format and displayed on the map without loss of precision or accuracy*.

- *The projection used for the ancillary layers is to match that used for the facility locations*. For example, if the ancillary layers (states and counties) are in North American Datum of 1927, but the facility latitude/longitude coordinates are in the North American Datum of 1983, there will be inaccuracy in the location of the facilities as it relates to the boundaries. Facilities near state boundaries could appear to be in the wrong state.

- *Only those facilities of interest in the spreadsheet are to be displayed on the map*.

- *Facilities that are not in the continental United States (for example, Guam, Hawaii, Alaska, etc.) need to be considered*. That is, make sure the requestor has specified whether they are to be shown or not.

Appendix C may provide additional information that would be useful when deciding what types of quality characteristics may be considered and documented in the Quality Objectives and Criteria (A7) element.

Special Training/Certification (A8)**—**In this example, the geospatial professional has the required background and experience to perform the work. However, if the map product were to be used in an official EPA publication, requirements for cartographic training might need to be specified here.

4.1.2 Group B: Measurement/Data Acquisition

The first eight elements addressing sampling and measurements are not required in this project because no new data collection will take place. These elements may be included in the QA Project Plan with the text "Not Applicable" next to each.

Data Acquisition Requirements (Nondirect Measurements) (B9)**—**Describe the sources of each data set to be used in the map. For example, describe or document

- the name of the individual who provided the spreadsheet (if different than the requestor);
- when the spreadsheet was delivered;
- the format (program) of the spreadsheet;
- the origin of the spreadsheet (it is very important to know where the requestor got the facility locations. The requestor is presumably NOT the originator of the latitude/longitudes but was provided them from some other source.)
- existing information about how the facility locations were derived;
- the format of the latitude/longitude coordinates;

- the date the locations were derived (does the date the locations were acquired affect the purpose of the map? For example, if the locations were derived ten years ago, but the map is to show the current set of facilities, there may be increased uncertainty as to the accuracy of the data.);
- the contents or metadata for the other data layers on the map.

Data Management (B10)—Describe how the data will be managed once acquired from the requestor. For a small project such as this, consider the following:

- Describe the applications format to be used to store the converted spreadsheet data file (e.g., dBase, Microsoft Access, INFO, other).
- Document any changes to field definitions necessary when converting the spreadsheet.
- Document the computer path to the data file(s) along with the names of the original input file and the names of any files created during the process of converting the data to GIS format.
- Document the input and output projection parameters used to reproject the data into a map-based coordinate system.
- Document and describe any custom subprograms used to process the data or to create the map.
- Describe the GIS software programs and versions used to process the data.

4.1.3 Group C: Assessment/Oversight

Assessments and Response Actions (C1)—The primary assessments to be described for this project would include:

- the method of ensuring that all spreadsheet records were properly translated into GIS records, including codes, numbers, and records (describe how the GIS data will be assessed to ensure that data were transferred correctly);
- the method of ensuring that the resulting map accurately shows the locations of the entities from the spreadsheet;
- the method of ensuring that there are no errors (typos, missing elements) in the map itself;
- the method of correcting errors found during the assessment.

Reports to Management (C2)—For this project, reports to management may only be required at the end of the task. In the Reports to Management (C2) element, discuss the content and scope expected in the final report. The final report may simply be an e-mail or informal memorandum, describing the completion of the project, the map deliverables, any problems encountered and their resolution.

4.1.4 Group D: Data Validation and Usability

Data Review, Verification, and Validation (D1)—State the criteria used to review and validate—that is, accept, reject, or qualify—data in an objective and consistent manner.

In a narrow scope project like this one, it may be difficult to objectively state criteria the data need to meet. It may be more appropriate to explore the data quality and report to the map requestor any omissions, problems, or concerns with the data.

Verification and Validation Methods (D2)—Describe the process for validating and verifying the data. Describe how the results will be communicated. In a project like this, the input data would be explored in an informal fashion to locate any problems. Some examples of data exploration include the following:

- Does every facility contain a latitude/longitude coordinate? List those that do not.

- Are the latitude/longitude coordinates consistent in their precision? For example, do some records contain data only to whole degrees while others contain more precise latitude/longitudes? If so, is there a question about variability in the quality of the data?

- Do the latitude coordinates contain leading "-" (minus signs) indicating locations in the Western hemisphere? Are all of the records consistent with regard to the use of minus signs for longitude?

- Do there appear to be any transpositions of latitude/longitude in the file? Create a simple map of the latitude/longitude coordinates. Do any of them appear in strange locations (for example, far outside the continental U.S.)?

Reconciliation with User Requirements (D3)—A limited scope project like this one has probably not undergone an extensive, systematic planning process. Therefore, this element can be used to communicate any potential problems found with the data file when compared to the performance criteria provided for the intended use.

After reviewing the input data set (as above), create a summary for the requestor indicating the nature of any omissions, errors, questions, or concerns about the data and their impact on the intended use. It is important to note that in a project like this, the requestor may not have personally reviewed the data and, therefore, may not be aware of potential problems. By providing a summary report, the requestor is given the option of modifying the map request, seeking clarification from the data originator on questions, and/or withdrawing the request.

4.2 MEDIUM DOCUMENTATION EXAMPLE: ROUTINE GLOBAL POSITIONING SURVEY TASK TO PRODUCE A GIS DATA SET

The example illustrates how elements B1 through B8 would be used when collecting primary geospatial data. The other two graded-approach examples concentrate more on the Data Acquisition Requirements (Nondirect Measurements) (B9) and Data Management (B10) elements issues related to the use of existing data rather than on the approaches used for new data specifically collected for a particular project.

A QA Project Plan for this task would document task-specific objectives for the survey and data evaluation criteria for the locational data to be collected. The task description and roles and responsibilities would be related to standard operating procedures and reporting forms of a single organization to avoid redundancy of documentation. Evaluation tasks would be specified to produce reports needed for product acceptance (or rejection). If accepted, “truth in labeling” information for the data set would be reported as standard metadata and entered into the GIS.

An adequate level of detail would be needed to clearly communicate agreed-upon survey objectives, data quality indicator criteria, and assessment and reporting requirements.

4.2.1 Group A: Project Management and Systematic Planning to Define the Task

The project management elements would emphasize task roles and responsibilities for planning and documenting the objectives of the task, evaluation criteria, and required assessments. Requirements for metadata records would also be documented.

Distribution List (A3)—The distribution list for the QA Project Plan on a project like this might include the EPA QA Officer, the EPA Task Leader, EPA Project Manager, GIS analysts, GPS technicians, and field staff.

Project/Task Organization (A4)—This element might describe the roles and responsibilities of each team member and provide an organization chart illustrating lines of communication and chain-of-command responsibilities. The organization description would clearly identify individuals with responsibility for developing, reviewing, and approving the QA Project Plan. Roles and responsibilities would be defined for field data collection, data management and processing, data quality assessment, reporting to the user, and records management.

Problem Definition/Background (A5)—The problem definition and background statement would describe the regulatory or decision-making context in which the project is operating. For example, describe the driving force behind the data collection effort and describe how the data will be used and by whom.

Project/Task Description (A6)—In this example project, the description would clearly state that the project will collect precise latitude/longitude coordinates using GPS equipment and that the results of the data collection process will be a complete and accurate GIS database of these locations, along with descriptive attributes. The project involves fieldwork and the use of GPS measurement equipment; therefore, the project/task description could discuss the basic assumptions and environment in which the project will utilize these methods.

Quality Objectives and Criteria (A7)—The user would provide criteria for acceptable data quality indicators such as accuracy (e.g., consistent with the EPA Locational Data Policy and standards), equipment sensitivity, precision, comparability, and completeness. Language from standard operating procedures could be used to describe the data quality requirements and to specify the criteria by which the collected data would be assessed.

Special Training/Certification (A8)—Describe how the field staff will be trained on the proper use of the GPS receiver, if necessary.

Documentation and Records (A9) —Requirements for task record keeping and/or metadata specifications or standards (e.g., EPA Method Accuracy and Description Codes or Federal Geographic Data Committee Standards) would be documented or included by reference. A data dictionary might also be described to fully document the database column names, types, widths, and contents, including any numeric coding schemes used to store nonlocational (attribute) data.

4.2.2 Group B: Data Collection

The data collection elements would describe in detail the implementation of standard operating procedures for field data collection (included by reference) during data management. The hardware/software configuration would be briefly described to document planned requirements and appropriate standard operating procedures to assure usefulness of the data set.

Sampling Process Design (B1)—To meet the task objectives and data quality indicator criteria developed in systematic planning, a survey design would be developed describing the sampling targets, sampling time, and frequency of data collection. Documentation of the design would include the rationale for choosing the specific sites to be sampled.

Sampling and Image Acquisition Methods (B2)—This element would be used to describe the actual procedures and methods used to collect the locational data using the GPS devices. Existing standard operating procedures such as those developed in EPA Region 5 and EPA Region 8 for GPS data collection could be cited or referenced, if those procedures will be used on this project. Include any special considerations regarding property access, transportation, or other logistical issues in this element.

Sample Handling and Custody (B3)—GPS data collection results in electronic files that will be downloaded and processed using GIS software. Therefore, there is no physical sample handling. This element might be used to describe how the electronic files from the GPS receivers are to be transmitted to the processing computers and who will do the transmitting.

Quality Control (B5)—Describe the overall quality control methods used to ensure that the locations for which latitude/longitude coordinates are collected meet the sampling design and are of the quality as set forth in the Quality Objectives and Criteria (A7) element.

Identify QC activities and the method to be used to obtain measurements. Describe the corrective action if the measurement is outside the performance limits.

Establish quality control methods for key entry, digitizing, or manually entering data to make ensure the data are correct. For example, provide a checklist to make sure field staff stand over the correct locations for the specified amount of time for GPS measurements. Measurements and observations can be compared to standard measurements and observations, or assessed against tolerance limits, to determine whether the data collection equipment is functioning within acceptable bounds or performance limits. Specify performance measures, measurement methods used, and the acceptable performance limits.

Instrument/Equipment Testing, Inspection, and Maintenance Requirements (B6)—Describe the procedures to be used to test, inspect, and maintain the GPS receivers. If standard operating procedures will be followed, cite them rather than duplicating their content here.

Instrument Calibration and Frequency (B7)—Note when periodic calibration of GPS equipment is to be performed. Describe the method of calibration and the frequency. Also, note where the calibration results are to be documented so they can be assessed before each GPS receiver is checked out for use. Cite—rather than reproducing—existing calibration procedures already specified in existing GPS standard operating procedures.

Inspection/Acceptance Requirements for Supplies and Consumables (B8)—Include a requirement to check batteries for the GPS receivers before commencing fieldwork. Discuss the requirement that batteries for each GPS receiver be fully charged and that any backup batteries also be charged and ready to go prior to fieldwork.

Data Management (B10)—For this project, data management activities would involve the storage and conversion of the GPS coordinates and associated attributes into GIS format and the subsequent data processing and manipulations of the coordinate and attribute data necessary for the final database to meet requirements for content, accuracy, projection, and format. Describe the procedures to be used during these processing steps in order to provide a complete overview of data management and manipulation. Describe any file naming conventions to be followed.

4.2.3 Group C: Assessment and Oversight

These elements would focus on the activities for assessing the effectiveness of project implementation and associated QA and QC activities to ensure that the QA Project Plan and its standard operating procedures are implemented as prescribed, including reports to project management and their response actions.

Assessments and Response Actions (C1)—Performance evaluations subsequent to training would document any GPS operator problems. Readiness reviews would include checks on equipment function such as sensitivity of detection and precision, correct recording and processing menus, base station availability, and survey logistics. The individuals or organizational units who will perform the assessments would be designated (e.g., regional coordinator, task manager). Standardized checklists can be used. During the survey, quality control procedures would be performed such as checks for accuracy against benchmarks. Any deviations from the task data collection design (e.g., lack of property access, interference) would be noted during the daily verification of data collection and reported, as well as field observations in designated forms to meet reporting requirements (EPA Method, Accuracy, and Description code requirements).

Assessment and differential correction would be performed with the designated software and base or reference station information before processing to produce the input file. Data quality assessments would include checking final data point locations with the field map for completeness, verifying that data quality indicator criteria were met, that metadata are adequate, and that files were adequately transferred and backed up. Input files for the GIS would be checked by an independent reviewer (e.g., regional coordinator or task manager) to assure they were complete, adequately documented to controls, and that they meet data quality indicators such as sensitivity, accuracy, precision, completeness, and if appropriate, comparability.

As manipulation of the coordinate data in the GIS occurs, continued assessments of the quality and accuracy of the manipulations would take place to ensure no discrepancies were introduced as a result of processing errors. Describe these checks and assessments and note when they would be made during the process of generating the GIS data set.

Reports to Management (C2)—Describe appropriate feedback loops to project management (e.g., Regional Coordinator) to assure prompt corrective action (e.g., GPS unit repair).

4.2.4 Group D: Data Validation and Usability

Use Group D elements to describe how field notes, reports, and other documents would be used to verify and validate the measured locations. These elements would also be used to describe how the data will be verified and validated. These activities address the data quality assessments that occur after data are collected and downloaded to a personal computer.

Data Review, Verification, and Validation (D1)—Once the final data set has been created, it would be reviewed, verified, and validated to ensure that it satisfies the quality, accuracy, and completeness required as defined in the Quality Objectives and Criteria (A7) element. Describe this review process in the Data Review, Verification, and Validation (D1) element. Describe what will be reviewed, verified, and validated.

Verification and Validation Methods (D2)—Describe how the final GIS data set will be validated and reviewed. For example, describe how the final data set's attribute tables will be compared to the data dictionary [as specified in the Documents and Records (A9) element] to ensure that the format and content of the data files are correct. Also describe how the locations of the final data set will be compared to both the original locations collected by the GPS receivers and the actual, true locations of the features collected. Verify that the EPA Method, Accuracy, and Description codes are present and accurately reflect the data collection process.

Reconciliation with User Requirements (D3)—This element would describe how the results of the data assessments, validations, and verifications will be compared and reconciled with criteria developed to ensure that the final deliverables (geospatial data or nongeospatial data files) are of sufficient quality to satisfy project requirements. For this project, document whether the final data meet, do not meet, or partially meet the quality objectives set out in the Quality Objectives and Criteria (A7) element. This might include descriptions of the success in capturing all the desired locations, noting whether postprocessing of the GPS coordinates resulted in sufficient locational accuracy, as specified in the Quality Objectives and Criteria (A7) element. If not, the impact on the intended use needs to be discussed.

4.3 COMPLEX DOCUMENTATION EXAMPLE: DEVELOPING COMPLEX DATA SETS IN A GIS FOR USE IN RISK ASSESSMENT MODELS

This project is to produce GIS database products that will be integrated into a risk assessment model. Risk assessment modelers and scientists would define the requirements for the geospatial products for their model in iterations with geospatial professionals. This project would involve digitizing spatial data sets from map sources, acquiring and converting existing data, creating subprograms within commercial off-the-shelf software to generate data, performing spatial analyses between GIS layers (for example, using spatial overlays to compare land use and demographic data), creating GIS databases for use in risk assessment models, and creating maps. The project would also involve interactions with risk assessment modelers and scientists, who would describe the geospatial products required for their models.

4.3.1 Group A: Project Management

Title and Approval Sheet (A1)—The approval sheet would include individuals who will define the GIS input data requirements for the models, accept the GIS data prior to inclusion in the models, review and check the geospatial data against the acceptability requirements, and check the subprograms created in the commercial off-the-shelf software to ensure they are working correctly. The project manager approving the project for implementation and the organization's QA Manager would also be included.

Distribution List (A3)—Provide names and addresses of participating project managers, QA Managers, and representatives from each technical team working on the project (planners, suppliers, and reviewers).

Project/Task Organization (A4)—Provide the participating project managers (client and supplier), QA Managers, and representatives from each technical team working on the project (planners, suppliers, and reviewers), listing their roles and responsibilities. An overall QA Project Plan created for the larger risk assessment modeling project might serve as a starting point for this element. The project organization chart and task descriptions can be expanded with information on the roles of those involved with the geospatial portion of the project.

Problem Definition/Background (A5)—Includes a summary definition of the problem, background of the overall project, as well as specific problem definitions and backgrounds of the geospatial portion of the project. One could summarize the Problem Definition/Background (A5) element of the QA Project Plan for the risk assessment modeling project as a whole, adding additional information relevant to the geospatial processing portion that is the focus of this QA Project Plan.

Project/Task Description (A6)—Focus on the project description and tasks for the geospatial processing project, integrating them with the schedule for the overall risk assessment project. The project/task description for the geospatial processing portion might include general descriptions of the data sources, processing steps, and data outputs to be created. Schedules would be defined, quality assessment techniques would be outlined, and quality assessment documentation and reports to the clients to be produced would be described.

Quality Objectives and Criteria (A7)—Establishing quality criteria for the information product output and relating it to data quality indicators to be checked within implementation of the data processing project is often difficult to do for geospatial projects of moderate to high complexity. In general, the data quality problems have much more to do with processing procedures (e.g., incorrect calculations, projections, programmatic manipulations, or procedural oversights) than with the ultimate locations of geographic entities to be analyzed or with source maps or data. Missteps in processing procedures often lead to nonsensical or incorrect data being produced or manipulated in future steps. Specific geospatial locations may be correct, but the attribute data produced for them may be incorrect.

If possible, state the requirements for positional accuracy. General qualitative statements are often the only possible way of describing the quality objectives for geospatial processing (e.g., *fuzzy tolerances used during processing will be set to the smallest possible level in order to ensure that data processing steps do not negatively affect existing locational accuracy*). Other examples of narrative descriptions of quality objectives include the following:

- *Reprojections, transformations, and other procedures that modify locational information must result in positional data that is accurate to the level of precision of the geospatial software being used.*

- *When digitizing data from map sources, be sure to document the acceptable root mean square error. This number is a measure of how closely the digitizer was able to match the source document to known geographic coordinates and, ultimately, is a measure of the positional accuracy achieved in converting paper maps into digital format.*

When performing attribute manipulations using database calculations, transformations, or formulas, it is presumed that no error is acceptable. Equations should be checked to assure they are coded correctly, and if they are, there are likely to be no errors in the resulting data. In other words, it would not make sense to say 90% of the resulting data are to be within 1% of the correct apportioned population.

Special Training/Certification (A8)—Any special training or experience in operating the commercial off-the-shelf software would be noted here.

Documentation and Records (A9)—Describe the requirements for documentation on the project. Policies for establishing metadata, especially a description of which FGDC-compliant metadata will be captured and how the metadata will be stored and managed would be included. Information on how the methodological procedures used on the project would be captured and documented might be included. For example, in geospatial projects where many steps are taken to configure, process, convert, transform, and manipulate the various data layers, taking careful note of procedures as they are developed is advantageous. This element could be used to specify how those notes will be entered into a document, at what level of detail, and how they will be used later in the project.

When subprograms written in commercial off-the-shelf software environments are to be developed, this element would be used to specify requirements for internal documentation of subprograms (e.g., program header information, and requirements for in-line program comments), and for external documentation of subprograms (e.g., summaries of the subprogram's purpose, inputs, outputs, and functions).

4.3.2 Group B: Measurement/Data Acquisition

Sampling Process Design (B1)—In this project, all of the marked-up maps provided by the survey respondents are to be digitized and entered into the GIS. Therefore, the Sampling Process (B1) element would simply state this requirement.

Sampling and Image Acquisition Methods (B2)—Since 100% of the source maps will be entered into the GIS, this element might simply state that this is a 100% sample.

Sample Handling and Custody (B3)—As part of this project, one or more maps will be received from industrial sites, indicating the location of their facilities and related features of interest (e.g., wells, property boundaries, and other information). These maps serve as source material and are to be handled and managed very carefully. This element would be used to describe any procedures for storing the maps, managing a check-in and check-out procedure so that each map's whereabouts are known, and documenting how these source materials will be handled so that none are lost or damaged.

Quality Control (B5)—Quality control procedures for the digitizing process would be documented in this element. These include procedures that indicate exactly how each map will be registered to the digitizing table, which features will be digitized, how features will be given identifying codes, and, especially, how at the completion of digitizing each resulting GIS data set will be checked against the original map to ensure that all required features have been digitized correctly.

Instrument/Equipment Testing, Inspection, and Maintenance Requirements (B6)—Document any inspection, testing, or maintenance recently performed or required to ensure that the digitizing table (or tables) are operating within the vendor's specified tolerance.

Instrument Calibration or Standardization and Frequency (B7)—Occasionally, digitizing tables will encounter calibration or operation problems causing incorrect or erroneous coordinates to be captured. Describe any calibration procedures (usually obtained from the manufacturer) that will be used to ensure that the precision of the digitizer is within specifications provided by the vendor.

Data Acquisition Requirements (Nondirect Measurements) (B9)—Describe the sources of each data set to be used in the project as follows:

- Define the source of each data layer to be used. Include the metadata provided with each layer. Some of the most important metadata elements include source citation, source scale, date of production, completeness, and use restrictions.

- Describe how each source will be used during the project.

- Describe why each existing data source was chosen for use in the project. What are the reasons these particular data sets are deemed to be superior to others (if more than one option exists)?

- Describe checks to be performed on the existing data to ensure that they were generated correctly and have the predicted content, format, and projection.

- For existing data received from unknown sources (e.g., spreadsheet data provided by other team members), quality checks would be extensive. Describe these checks (logical consistency, completeness, geospatial location accuracy, etc.).

Data Management (B10)—Describe how the data will be managed once acquired from the requestor. For this complex task, the Data Management (B10) element will be quite extensive, including information on the following topics:

- path names to all data sources to be used on the project;

- methods to be employed to ensure that any informal subprograms will be developed and tested to ensure they operate as expected (e.g., accurate calculations);

- a description of the formats of the data sources, any interim or temporary data sets to be created, and the final data products;

- a data dictionary that describes, for each source database and the final product, the content, type, name, and field width of each attribute;

- if a full requirements-design-development-testing process is to be carried out for any programs to be written, documentation of that development process, including the documents that resulted from that process in the Data Management (B10) element.

4.3.3 Group C: Assessment/Oversight

Assessments and Response Actions (C1)—At each processing step on this project, quality assessments are to be performed to ensure that the data sources, interim products, and final databases meet quality objectives. In the Assessments and Response Actions (C1) element, include methods for ensuring that:

- all source maps were digitized;
- all source features were accurately digitized;
- each map source was registered to within specified tolerances on the digitizing tablet (creating checklists to track these assessments might be helpful);
- attribute codes and categorical data assigned to digitized features were complete and accurate;
- each existing data source used was downloaded completely and without corruption of coordinates or attributes;
- each existing data source has the correct input coordinate system information
- any reprojections/transformations of input data sets were carried out correctly (including datum shifts, if applicable);
- each processing step or "macro" was performed correctly and was performed on the correct input data;
- proper coordinate precision (e.g., single precision or double precision) was maintained throughout each step of the process;
- there was no unacceptable loss of precision or rounding of coordinates throughout processing due to raster-to-vector conversions, topological rebuilds, or other procedures;
- calculations resulting in new data fields are performed correctly, that any constants used were entered correctly, and that the resulting data are within expected ranges.

For each of the assessment methods above, describe the methods to be used to correct the problem and reprocess any resulting data sets.

Reports to Management (C2)—Describe the interim reports to be submitted to management throughout the project and note the frequency and content expected for each. For this risk assessment project, reports to management might include

- weekly or biweekly reports describing progress, problems, errors encountered, or unexpected occurrences;

- monthly summary of processing status (Which data layers have been processed and through which stage of the project? Include information about any sites that require special processing. For example, if there are any sites outside the continental United States, what special provisions for coordinates systems, projections, and precision need to be made?);

- final reports indicating overall processing results, identifying the products created, and describing the assessment methods used to gauge accuracy [use information from the Assessments and Response Actions (C1) element].

4.3.4 Group D: Data Validation and Usability

In most geospatial projects, the Group D elements will describe the process of checking and validating the final data or maps to be delivered. If the activities in the Group C elements are properly carried out during the course of the project, the Group D elements will uncover few problems.

Data Review, Verification, and Validation (D1)—State the criteria used to review and validate—that is, accept, reject, or qualify—data in an objective and consistent manner. For this project, this element would include a description of the criteria used to assess whether the final deliverables are correct. For this project, any errors, omissions, corrupted data files, incorrect calculations, or missing information would result in rejection and reprocessing of the final files. It is hoped that any errors detected in the final data files or coverages are the result of problems in the last stages of processing. This assumes that the actions carried out in Group C have identified errors and problems during early and middle stages of production.

Verification and Validation Methods (D2)—Describe the process for validating and verifying the data and how the results will be communicated. In addition, for this element, describe:

- the method for reviewing each final data set to be delivered, in general terms;

- specific methods for reviewing each data set [For example, if the data sets to be delivered are a set of database files containing such things as the populations for each land-use type within a certain distance of an industrial facility, this element would be used to describe checks to ensure that the final data files contain the appropriate numbers of records (e.g., all of the census block groups over the entire study area are accounted for) and that the population aggregations or disaggregations have been done correctly (e.g., there are no negative population counts and spot checks indicate that population summaries are correct by performing manual calculations];

- the method for ensuring that each data file has been not corrupted and can be uncompressed (if compressed for delivery).

If the actions described in Group C are followed, any problems encountered at this stage would be limited to the generation of the final deliverable files themselves—not to a serious flaw in the methodology or steps performed earlier in the project.

Reconciliation with User Requirements (D3)—This element would describe how the results of the data assessments, validations, and verifications will be compared and reconciled with criteria developed to ensure that the final deliverables (geospatial or nongeospatial data files) are of sufficient quality to satisfy project requirements. For this project, this element would document whether each component of the final deliverables (i.e., each data file or spatial data layer) meets, does not meet, or partially meets the quality objectives stated in the Quality Objectives and Criteria (A7) element. For example, did all database calculations that created new database fields produce correct results? When comparing the spatial locations of lines and polygons in final output data sets to original data sets, was there any inappropriate movement of those features? If there were problems, errors, or inconsistencies, the Reconciliation with User Requirements (D3) element would include a description of how these problems will affect usability of the final data sets.

APPENDIX A

BIBLIOGRAPHY

American National Standards Institute/American Society for Quality Control (ANSI/ASQC). (1995). *Specifications and Guidelines for Quality Systems for Environmental Data Collection and Environmental Technology Programs (E4-1994)*. American National Standard.

Federal Geographic Data Committee. (1997). Content Standards for Digital Geospatial Metadata, Federal Geographic Data Committee, Washington, DC.

Institute of Electrical and Electronics Engineers (IEEE). (1998). Standard 830: IEEE Recommended Practice for Software Requirements Specifications. IEEE Standards Collection: Software Engineering (Volume 4: Resource and Technique Standards). Piscataway, NJ.

National Institute of Standards and Technology (NIST). (1994). *Federal Information Processing Standards Publication 173-1*. Gaithersburg, MD. Available: *http://www.itl.nist.gov/fipspubs/.*

U.S. Environmental Protection Agency. (1998a). *EPA Guidance for Quality Assurance Project Plans (QA/G-5)* (EPA/600/R-98/018). Washington, DC: Office of Research and Development.

U.S. Environmental Protection Agency. (1998b). *Information Resources Management Policy Manual* (Directive 2100). Washington, DC.

U.S. Environmental Protection Agency. (2000a). *EPA Quality Manual for Environmental Programs* (Order 5360 A1). Washington, DC.

U.S. Environmental Protection Agency. (2000b). *Guidance for Data Quality Assessment: Practical Methods for Data Analysis (QA/G-9)* (EPA/600/R-96/084, QA00 Update). Washington, DC: Office of Environmental Information.

U.S. Environmental Protection Agency. (2000c). *Guidance for the Data Quality Objectives Process (QA/G-4)* (EPA/600/R-96/055). Washington, DC: Office of Environmental Information.

U.S. Environmental Protection Agency. (2000d). *Policy and Program Requirements for the Mandatory Agency-wide Quality System* (EPA Order 5360.1 A2). Washington, DC.

U.S. Environmental Protection Agency. (2001a). *Geospatial Baseline Report*, Office of Environmental Information, Washington, DC.

U.S. Environmental Protection Agency. (2001b). *EPA Requirements for Quality Assurance Project Plans (QA/R-5)* (EPA/240/B-01/003). Washington, DC: Office of Environmental Information.

Veregin, H. (1992). *GIS Data Quality Assessment Tools*. Internal research project report. Las Vegas, NV: Environmental Monitoring Systems Laboratory, U.S. Environmental Protection Agency.

APPENDIX B

GLOSSARY

acceptance criteria – specific limits placed on an item, process, or service defined in requirements documents. Acceptance criteria are acceptable thresholds or goals for data, usually based on individual data quality indicators (precision, accuracy, representativeness, comparability, completeness, and sensitivity).

accuracy – the degree to which a calculation, measurement, or set of measurements agree with a true value or an accepted reference value. Accuracy includes a combination of random error (precision) and systematic error (bias) components which are due to sampling and analytical operations. A data quality indicator. EPA recommends that this term not be used and that precision and bias be used to convey the information usually associated with accuracy.

address geocoding – assigning x,y coordinates to tabular data such as street addresses.

attribute – any property, quality, or characteristic of sampling unit. The indicators and other measures used to characterize a sampling site or resource unit are representations of the attributes of that unit or site. A characteristic of a map feature (point, line, or polygon) described by numbers or text; for example, attributes of a tree represented by a point might include height and species. (See related: *Continuous*)

attribute accuracy – the closeness of attribute values (characteristic of the location) to their true value, which includes continuous attributes with measurement error (e.g., elevation) and categorical accuracy resulting from misclassification (e.g., soil types on a soil map).

band – one layer of a multispectral image that represents data values for a specific range of reflected light or heat—such as ultraviolet, blue, green, red, infrared, or radar—or other values derived by manipulating the original image bands.

bias – in a sampling context, the difference between the conceptual, weighted average value of an estimator over all possible samples and the true value of the quantity being estimated. An estimator is said to be unbiased if that difference is zero. The systematic or persistent distortion of a measurement process that deprives the result of representativeness (i.e., the expected sample measurement is different than the sample's true value). A data quality indicator.

cell size – the area on the ground covered by a single pixel in an image, measured in map units.

classification – the process of assigning a resource unit to one of a set of classes defined by values of specified attributes. For example, forest sites will be classified into the designated forest types, depending on the species composition of the forest. Systematic arrangement of objects into groups or categories according to established criteria

comparability – the degree to which different methods, data sets, and/or decisions agree or can be represented as similar.

completeness – the amount of valid data obtained compared to the planned amount, usually expressed as a percentage.

computer-aided design package – an automated system for the design, drafting, and display of graphical information.

continuous – a characteristic of an attribute that is conceptualized as a surface over some region. Examples are certain attributes of a resource, such as chemical stressor indicators measured in estuaries.

coordinates – linear and/or angular quantities that designate the position of a point in relation to a given reference frame.

Data Quality Indicators – quantitative and qualitative measures of principal quality attributes, including precision, accuracy, representativeness, comparability, completeness, and sensitivity.

Data Quality Objectives – qualitative and quantitative statements derived from the DQO Process that clarify study objectives, define the appropriate type of data, and specify tolerable levels of potential decision errors that will be used as the basis for establishing the quality and quantity of data needed to support decisions.

Data Quality Objectives Process – a systematic tool to facilitate the planning of environmental data collection activities. Data quality objectives are the qualitative and quantitative outputs from the DQO Process.

datum (plural datums) – in surveying, a reference system for computing or correlating the results of surveys. There are two principal types of datums: vertical and horizontal. A vertical datum is a level surface to which heights are referred. In the United States, the generally adopted vertical datum for leveling operations is the National Geodetic Vertical Datum of 1929 (see below). The horizontal datum is used as a reference for position. The North American Datum of 1927 (see below) is defined by the latitude and longitude of an initial point (Meade's Ranch in Kansas), the direction of a line between this point and a specified second point, and two dimensions that define the spheroid. The new North

American Datum of 1983 (see below) is based on a newly defined spheroid (GRS80); it is an Earth-centered datum having no initial point or initial direction.

digital elevation model – the representation of continuous elevation values over a topographic surface by a regular array of z-values, referenced to a common datum. Typically used to represent terrain relief.

digital line graph – digital data produced by the U.S. Geological Survey. These data include digital information from the U.S. Geological Survey map base categories such as transportation, hydrography, contours, and public land survey boundaries.

digital orthophotography – see *orthophotograph*y

digitizing table – an electronic device consisting of a flat surface and a handheld cursor that converts positions on the table to digital x,y coordinates.

feature – an entity in a spatial data layer, such as a point, line, or polygon, that represents a geographic object.

Federal Geographic Data Committee – the Federal Geographic Data Committee coordinates the development of the National Spatial Data Infrastructure (NSDI). The NSDI encompasses policies, standards, and procedures for organizations to cooperatively produce and share geographic data. The 17 federal agencies that make up the FGDC are developing the NSDI in cooperation with organizations from state, local, and tribal governments, the academic community, and the private sector.

Federal Information Processing Standard – standards approved by the Secretary of Commerce under the Information Technology Management Reform Act (Public Law 104-106). These standards and guidelines are issued by the National Institute of Standards and Technology (NIST) as Federal Information Processing Standards (FIPS) for use government-wide. FIPS coding standards include, for example, two-digit numeric codes used to identify each of the 50 U.S. states and three-digit numeric codes used to identify each U.S. county.

geographic feature – see *feature*.

Geographic Information System – a collection of computer hardware, software, and geographic data designed to capture, store, update, manipulate, analyze, and display geographically referenced data.

geospatial data – the information that identifies the geographic location and characteristics of natural or constructed features and boundaries on the earth. This information may be derived from, among other things, remote-sensing, mapping, and surveying technologies.

Global Positioning System – a constellation of 24 satellites, developed by the U.S. Department of Defense, that orbit the Earth at an altitude of 20,200 kilometers. These satellites transmit signals that allow a GPS receiver anywhere on Earth to calculate its own location. The Global Positioning System is used in navigation, mapping, surveying, and other applications where precise positioning is necessary.

graded approach – the process of basing the level of managerial controls on the item or work according to the intended use of the results and the degree of confidence needed in the quality of the results.

grid – a data structure commonly used to represent map features. A cellular-based data structure composed of cells or pixels arranged in rows and columns (also called a *raster*).

ground-truthing – the use of a ground survey to confirm the findings of an aerial survey or to calibrate quantitative aerial or satellite observations.

imagery – visible representation of objects and/or phenomena as sensed or detected by cameras, infrared, and multispectral scanners, radar, and photometers. Recording may be on photographic emulsion (directly, as in a camera, or indirectly, after being first recorded on magnetic tape as an electrical signal) or on magnetic tape for subsequent conversion and display on a cathode ray tube.

kriging – a weighted, moving-average estimation technique based on geostatistics that uses the spatial correlation of point measurements to estimate values at adjacent, unmeasured points. A sophisticated technique for filling in missing data values, kriging is named after a South African engineer, D.G. Krige, who first developed the method. The kriging routine preserves known data values, estimates missing data values, and estimates the variance at every missing data location. After kriging, the filled matrix contains the best possible estimate of the missing data values, in the sense that the variance has been minimized.

landsat – a series of orbiting satellites used to acquire remotely sensed images of Earth's land surface and surrounding coastal regions.

leaf on/leaf off – the characteristic of deciduous vegetation based on seasonality. Refers to whether deciduous trees have leaves during image acquisition.

locational – of or referring to the geographic position of a feature.

Map Digitization – conversion of map data from graphic to digital form.

map projection – a mathematical formula or algorithm for translating the coordinates of features on the surface of the Earth to a plane for representation on a flat map.

map resolution – the accuracy with which the location and shape of map features are depicted for a given map scale.

map scale – a statement of a measure on the map and the equivalent measure on the Earth, often expressed as a representative fraction of distance, such as 1:24,000.

map, thematic – map designed to provide information on a single topic, such as geology, rainfall, or population.

metadata – information about a data set. Metadata for geographical data may include the source of the data; its creation date and format; its projection, scale, resolution, and accuracy; and its reliability with regard to some standard.

Method, **Accuracy, and Description Data –** a coding scheme developed by EPA to promulgate standards for describing the type and quality of spatial data. The coding scheme includes both database field definitions and standardized codes.

modeling – development of a mathematical or physical representation of a system or theory that accounts for all or some of its known properties. Models are often used to test the effect of changes of components on the overall performance of the system.

National Geodetic Vertical Datum of 1929 – reference surface established by the U.S. Coast and Geodetic Survey in 1929 as the datum to which relief features and elevation data are referenced in the conterminous United States; formerly called "mean sea level 1929."

National Hydrography Data Set – a comprehensive set of digital spatial data that contains information about surface water features such as lakes, ponds, streams, rivers, springs, and wells.

National Map Accuracy Standards – specifications promulgated by the U.S. Office of Management and Budget to govern accuracy of topographic and other maps produced by federal agencies.

National Institute of Standards and Technology – a non-regulatory federal agency within the U.S. Commerce Department's Technology Administration whose mission is to develop and promote measurement, standards, and technology to enhance productivity, facilitate trade, and improve the quality of life. NIST laboratories provide technical leadership for vital components of the Nation's technology infrastructure needed by U.S. industry to continually improve its products and services .

National Land Cover Data – a nationally consistent land-cover data set developed by the National Land Cover Characterization program.

National Spatial Data Infrastructure – the technologies, policies, and people necessary to promote sharing of geospatial data throughout all levels of government, the private and nonprofit sectors, and the academic community. The NSDI was established in 1994 by Executive Order 12906.

North American Datum of 1927 – the primary local geodetic datum used to map the United States during the middle part of the 20^{th} century, reference to the Clarke spheroid of 1866 and an initial point at Meade's Ranch, Kansas. Features on U.S. Geological Survey topographic maps, including the corners of 7.6-minute quadrangle maps, are referenced to this datum. It is gradually being replaced by the North American Datum of 1983.

North American Datum of 1983 – a geocentric datum based on the Geodetic Reference System 1980 ellipsoid (GRS80). Its measurements are obtained from both terrestrial and satellite data.

orthophotography – perspective aerial photography from which distortions owing to camera tilt and ground relief have been removed. Orthophotography has the same scale throughout and can be used as a map.

performance criteria – measures of data quality that are used to judge the adequacy of collected information that is new or original, otherwise known as "primary data."

photogrammetry – science or art of obtaining reliable measurements or information from photographs or other sensing systems.

positional accuracy – the closeness of locational information to its true position.

precision – (i) the degree to which replicate measurements of the same attribute agree or are exact. Precision is the degree to which a set of observations or measurements of the same property, usually obtained under similar conditions, conform to themselves. A data quality indicator (See related : *Accuracy, Bias*). (ii) The number of significant decimal places used to store floating point numbers (e.g., coordinates) in a computer. Single precision denotes use of up to seven significant digits to store floating point numbers. Double precision denotes use of up to 14 significant digits to store floating point numbers.

projection – a mathematical model that transforms the locations of features on the Earth's surface to locations on a two-dimensional surface.

quality assurance project plan – a document describing in detail the necessary quality assurance, quality control, and other technical activities that should be implemented to ensure the results of the work performed will satisfy the stated performance criteria.

quality assurance– an integrated system of management activities involving planning, implementation, documentation, assessment, reporting, and quality improvement to ensure that a process, item, or service is of the type and quality needed and expected by the client.

quality control – the overall system of technical activities that measure the attributes and performance of a process, item, or service against defined standards to verify that they meet the stated requirements established by the customer; also, operational techniques that are used to fulfill requirements for quality.

quality management plan – a document that describes a quality system in terms of the organizational structure, policy and procedures, functional responsibilities of management and staff, lines of authority, and required interfaces for those planning, implementing, documenting, and assessing all activities conducted.

raster data (raster image) – a spatial data model made of rows and columns of cells. Each cell contains an attribute value and location coordinates; the coordinates are contained in the order of the matrix, unlike a vector structure, which stores coordinates explicitly. Groups of cells that share the same value represent geographic features.

remote sensing – process of detecting and/or monitoring chemical or physical properties of an area by measuring its reflected and emitted radiation.

root mean square error – the square root of the average of the set of squared differences between dataset coordinate values and coordinate values from an independent source of higher accuracy for identical points.

representativeness – the degree to which data accurately and precisely represent the frequency distribution of a specific variable in the population.

scale – relationship existing between a distance on a map, chart, or photograph and the corresponding distance on the Earth.

Soil Survey Geographic (SSURGO) Data – a nationwide, geospatial, soils database created by the Natural Resources Conservation Service from 1:250,000-scale soil maps.

spheroid – an ellipsoid that approximates a sphere. Used to describe (approximately) the shape of the earth.

SSURGO – see *Soil Survey Geographic Data.*

tic – a point on a map representing a location whose coordinates are known in some system of ground measurement such as latitude and longitude.

topography – configuration (relief) of the land surface; the graphic delineation or portrayal of that configuration in map form, as by contour lines. In oceanography the term is applied to a surface such as the sea bottom or surface of given characteristics within the water mass.

Topologically Integrated Geographically Encoding and Referencing (TIGER) System – the data system developed by the U.S. Census Bureau to describe the boundaries of all census geography (e.g., states, counties, census tracts) and to tie decennial census tabulations to census boundaries.

topology – the spatial relationships between connecting or adjacent features in a geographic data layer. Topological relationships are used for spatial modeling operations that do not require coordinate information.

vector – a data structure used to represent linear geographic features. Features are made of ordered lists of x,y coordinates and represented by points, line, or polygons; points connect to become lines, and lines connect to become polygons. Attributes are associated with each feature.

APPENDIX C

SPATIAL DATA QUALITY INDICATORS FOR GEOSPATIAL DATA

The Federal Information Processing Standard (FIPS) 173 (NIST, 1994) emphasized five components of data quality that are basic to the Federal Geographic Data Committee metadata [Section 3.1.9, Records and Documentation (A9)]:

- Accuracy—positional
- Accuracy—attribute
- Completeness
- Logical consistency
- Lineage

In geospatial data, like that for other environmental data, accuracy is defined as the closeness of results to "true" values (surveying or remote-sensing reference points). All spatial data are inaccurate (have error) to some degree. Generally stated, error (r) is equivalent to the difference between the estimated value and the true value. Because a certain amount of inaccuracy is inherent in all locational measurements, the degree of inaccuracy must be assessed and compared to the accuracy required for the final geospatial data product.

There are two kinds of geospatial data accuracy:

- **Positional Accuracy** is the closeness of the locations of the geospatial features to their true position.

- **Attribute Accuracy** is the closeness of attribute values (characteristics at the location) to their true values. This applies to accuracy of continuous attributes such as elevation and accuracy of categorical attributes such as soil types.

Positional Accuracy

An example of the kinds of positional accuracy problems that may be encountered is illustrated in the map of Condea Vista, in southeastern Oklahoma City (Figure 1).

The polygon on the map represents the boundary of a Resource Conservation and Recovery Act site from a permit file map that was referenced to the U.S. Geological Survey 7.5-minute quad sheet and digitized. The points on the map are all estimates of the latitude/longitude of the site derived by various methods. Note the distribution of the points. All are valid, but some are not as accurate as others. Three points—ZIP code, PLSS, and an address match—fall outside the facility boundaries. In systematic planning, requirements for the project's positional accuracy need to be defined. Then, collected or acquired data are evaluated against those requirements. Reporting requirements for data providers or data producers document targets for accuracy (e.g., proof in labeling) and information for consumers to use in determining fitness for use. Accuracy targets such as the FGDC's National Standard for Spatial Data Accuracy Test Guidelines and EPA's Locational Reporting Standard of ± 25 meters might be referenced.

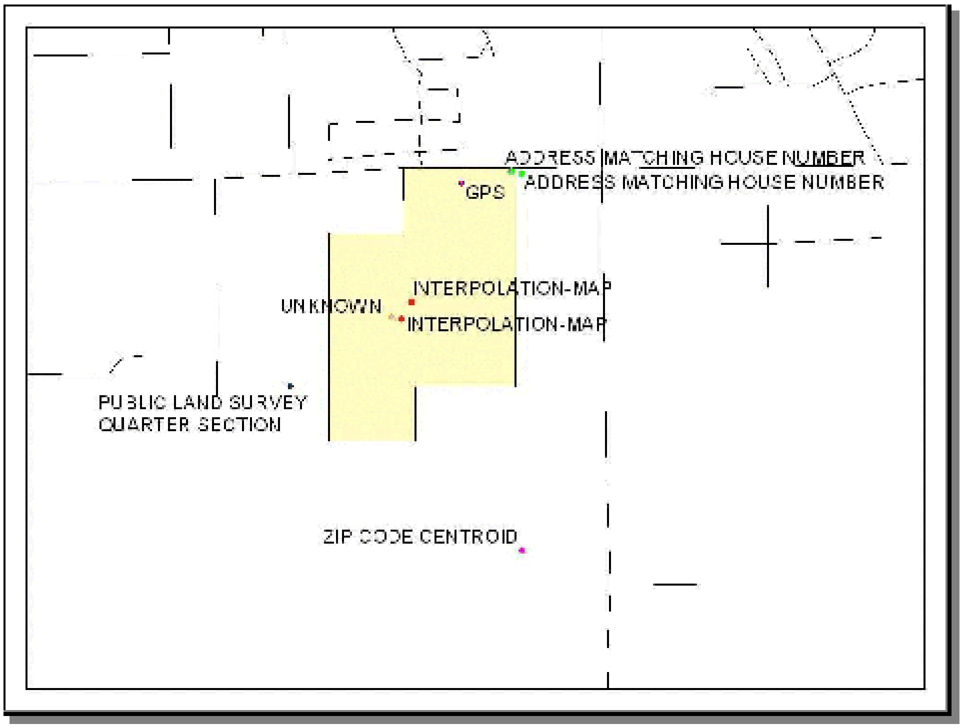

Figure 1. An illustration of how different methods of determining location result in different "answers." The method used to determine a facility location is a data quality indicator.

Accuracy can be assessed by comparing geospatial data to a source map or data of higher accuracy and determining statistical measures such as root mean square error and confidence levels (e.g., error bars on kriging contours) to judge the amount of inaccuracy. A rule of thumb is to use at least 20 points for comparison. For example:

- Evaluation Data Set: Envirofacts Address Matching Points
- Compared to higher accuracy source: Texas GPS border survey (20 points)
- Projection: National Lambert Meters, (North America Datum of 1983)
- Geographic area: Brownsville, TX to Las Cruces, NM
- Absolute difference in x range 8-669 m; y 8-1090 m
- Root mean square error (RMSE) (x) = 187; RMSE (y) = 257
- Accuracy = 2.4477*0.5*(RMSE(x) + RMSE(y)) = 544
- Reporting: Tested 544 meters horizontal accuracy at 95% confidence level

In systematic planning, it is important to set quality criteria for data or products being produced or for those acquired from another source such as a map or spatial data set. Determine the maximum error allowable in the product and see if it meets the project needs (e.g., EPA's target for location information is ±25 meters by GPS). The data producer may provide or be requested to provide statistics of accuracy for any acquired products. Identifying the steps used to produce or create the data set would be helpful in order to document any transformations between coordinate systems or reformatting that could impact accuracy. This could include estimating the error in each transformation or conversion and checking on the propagation of error between steps. For example, check the resolution of a product map by comparing the projection to known values and compute the root mean square error.

Attribute Accuracy

Attributes are facts tied to the Earth's surface. Attributes include qualitative facts like soil classification for areas of the Earth's surface on a soil map and quantitative facts like slope or population at a point on the Earth's surface. Attributes are linked to geographic features in a geospatial database via database identifiers. Attribute errors can be introduced from direct observation, remote-sensing interpretation, or interpolation and can affect the accuracy of the facts. Data producers need to provide accuracy information as proof of product.

For quantitative attribute accuracy, assessments can be carried out that vary with the data use and its complexity, such as

- assessing standard error for quantitative data (e.g., 7-meter uncertainty in slope value based upon known 1-meter standard deviation in elevation measurements)

- assessing or documenting known measurement error (e.g., Landsat "striping," where error exists in every 6th row in a scene and is removed by a simple arithmetic operation)

- development of uncertainty models and Monte Carlo analysis to determine uncertainty for spatial models.

For qualitative attributes accuracy, assessments can be carried out for classification of nominal errors. A standard must be identified for comparison of the evaluated data to "true" values such as ground-level observations of land characteristics, and the results reported for evaluation against an accuracy criteria such an error matrix. Such a standard and evaluation can provide the percentage of classification cases that are correct, percentage correctly classified, or a Kappa Index, which adjusts for correct identification by chance. As part of the systematic planning process, evaluation criteria (for example, accuracy or uncertainty criteria) need to be developed and used in evaluation of the data for fitness for use.

Completeness

Completeness is defined as the degree to which the entity objects and their attributes in a data set represent all entity instances of the abstract universe (defined by what is required for the project's data use in systematic planning). Metadata should provide a good definition of the abstract universe with defined criteria for selecting the features to include in the data set so the data user can perform an independent evaluation. Missing data (incompleteness) can affect logical consistency needed for correct processing of data by software.

Logical Consistency

A spatial data set is logically consistent when it complies with the structural characteristics of the data model and is compatible with attribute constraints defined for the system. In systematic planning, logical rules of structure (such as rules for topological relationships) could be identified, as well as rules for attribute consistency needed for appropriate data use. When acquiring data from another source or when creating new data, tests could be planned to check spatial data against those defined requirements. For example:

- In an electric utility application, a logical consistency rule may be in place indicating that electrical transformers must always occur on power poles. If so, ensure that each electrical transformer is assigned to a power pole. Those that are not are logically inconsistent.

- Are there valid attribute values for objects (e.g., for date attributes, the range of values must fall between 1 and 31, inclusive)?

Inconsistencies violate rules and constraints. Data should meet rules and constraints such as attribute range, geometric and topological constraints, and rules for spatial relationships in order to be used according to the project's requirements. Consistency is needed for control of transactions in database and software operations. Without consistency, additional time and effort must be expended to allow software to handle the inconsistencies in ways that do not propagate or increase the errors. Evaluations need to be reported in displays or written reports to characterize product quality.

Precision

Precision is a data quality indicator often used for environmental data that were, unfortunately, not included in the FIPS 173 list. It is defined as the number of decimal places or significant digits in a measurement (related to standard deviation around the mean of many measurements and rounding off). Although GIS software transactions are often more precise (more significant figures) than the data it processes, errors can occur (e.g., conversion of data with two significant figures, which displaces point locations to one, as shown in Figure 2). When the coordinates used to represent the locations of geographic features have low precision (that is, few significant digits), this might be an indicator of data quality that needs to be assessed. If the precision of the coordinates in the data are not sufficient to represent the geographic features to the degree required, this issue should be documented and a determination made as to whether the data will accommodate their intended use.

Lineage

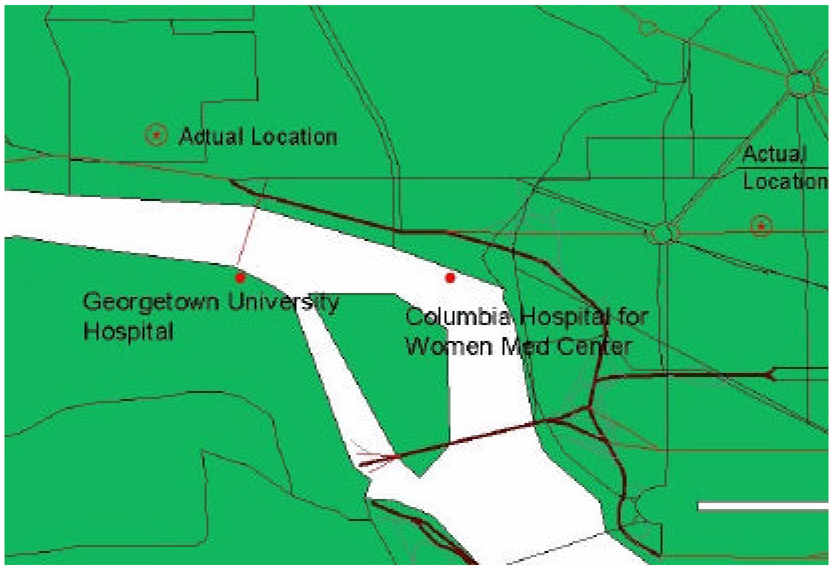

Figure 2. An illustration of how coordinates having less precision (for example, rounding up to the nearest degree of latitude/longitude), may not precisely reflect actual locations. Precision is a data quality indicator.

Data lineage is the description of the origin and processing history of a data set. It includes the name of the organization that produced the data so that its policies, procedures, and methods can be evaluated to see if they were biased in representing the surface of the Earth or its features. For example, if lineage indicates that the U.S. Geological Survey is the originator of a geospatial data set, then certain assumptions about their policies, procedures, and methods could be made. For example, the U.S. Geological Survey requires that no more than 10 percent of points tested on a map boundary can be in error by more than 1/30 of an inch at a scale of 1 inch to 20,000 feet. Lineage also provides references for data accuracy (for example, map accuracy standards), how accuracy was determined, and corrections made in producing the source map from which the data were derived. Lineage for general metadata provides spatial data quality characteristics such as accuracy, precision, and scale for a series of products. Information as to the coordinate systems used to reference locations (including necessary, unique projection parameters that are required to fully document map projections) are also components of lineage information in metadata.